ALGEBRA

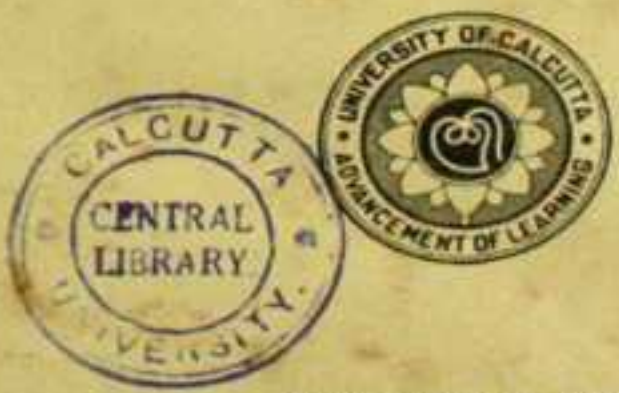
LECTURES DELIVERED TO POST-GRADUATE STUDENTS OF
CALCUTTA UNIVERSITY

BY

FRIEDRICH WILHELM LEVI, DR.PHIL.NAT.
HARDINGE PROFESSOR

Parts III - V

PART III—CONTINUED FRACTIONS
PART IV—APPROXIMATE SOLUTION
PART V—MATRICES. RESULTANTS



PUBLISHED BY THE
UNIVERSITY OF CALCUTTA
1937

PRINTED IN INDIA

PRINTED AND PUBLISHED BY BHUPENDRALAL BANERJEE
AT THE CALCUTTA UNIVERSITY PRESS, SENATE HOUSE, CALCUTTA.

Reg. No. 1062B—December, 1937—E

PREFACE

The third and the fourth part of these lectures deal with some classical portions of Algebra. Obviously a strict selection has been necessary, and the author gave preference to those portions of Continued Fractions (Part III) which are connected with the theory of numbers. In Part IV (Approximation solution) much importance has been given to how to carry out the calculations. Of course this branch of Algebra has to be considered from quite a different point of view from that of General Algebra. By approximation numerical values should be found out in the quickest and easiest way; so the reader cannot get a real insight into the nature and importance of the different methods without knowing the difficulties a practical reckoner has to face. The hints given by the author for the simplification of calculations do not involve the use of slide rules, calculating machines and graphical methods as these expedients are not familiar to our students. In this as well in some other items a later edition of these lectures is expected to show some alteration.

Part V contains the most important theorems on matrices with application on Hermitian and quadratic forms. Here the student may have the satisfaction of seeing in a nutshell the greater part of what he learned on Analytic Geometry.

As in the prefaces of Part I and of Part II, I have much pleasure in delivering my thanks here to friends and kind collaborators. The Vice-Chancellor of our University, Syamaprasad Mookerjee, Esq., M.A., B.L., M.L.A., Barrister-at-Law, gave me every possible help to get these lectures printed in a very short time, and is fully entitled to the grateful thanks of the author as well of the students. Proofs and manuscripts have been carefully revised by Mr. A. C. Chowdhury, M.Sc., Research Scholar of Calcutta University. The Calcutta University Press kindly co-operated in carrying out the printing in the proposed time.

F. W. Levi

Calcutta, Asutosh Building,
October, 1937

INDEX

PART III. CONTINUED FRACTIONS

PART IV. APPROXIMATE SOLUTION

PART V. MATRICES. RESULTANTS

PART III

CONTINUED FRACTIONS

§ 1. General properties of continued fractions.

Let K be a field, S a subring of K, and A be a subset of K with the following property: If a class of residues $\neq (0)$ of S contains an element of A, this class contains also the inverse of an element of A. Hence if [1/1]

$$a, a', a'', \ldots, a_1, a_2, \ldots \tag{1, 1}$$

denote elements of A and

$$s, s', s'', \ldots, s_1, s_2, \ldots \tag{1, 2}$$

denote elements of S, then every element of A can be represented either by

$$a = s + 1 : a' \tag{1, 3}$$

or by

$$a = s. \tag{1, 3'}$$

If *e. g.*, K is the field of the real numbers, S the ring of the integers, and A the set of the real numbers >1, the representation (1,3), (1,3') is always possible and s is uniquely defined by a as the greatest integer $\leqq a$.

But the conditions (1,3), (1,3') hold also for other sets A in fields K, so an investigation in general terms is helpful.

Let

$$\begin{aligned} a_1 &= s_1 + 1 : a_2 \\ a_2 &= s_2 + 1 : a_3 \\ &\ldots\ldots\ldots\ldots \\ a_n &= s_n + 1 : a_{n+1}, \end{aligned} \tag{1, 4}$$

then

$$a_1 = s_1 + \cfrac{1}{s_2 + \cfrac{1}{s_3 + \ddots + \cfrac{1}{s_n + \cfrac{1}{a_{n+1}}}}}, \tag{1, 5}$$

The representation of a_1 by (5) is said to be a *continued fraction*. The formulae (1,4) can be continued till an element a_i will be an element of S. If there is an m so that $a_m = s_m$, then the continued fraction is *finite*, otherwise it is *infinite*. If a can be represented by a finite continued fraction.

it belongs to the quotient field Q of S, and every finite set of elements

$$s_1, \ldots, s_n \tag{1, 6}$$

of S defines an element of Q by the help of (1,5).

We now define other sequences of elements of S by the following formulas.

$$P_{-1}=0,\ P_0=1,\ P_k=s_kP_{k-1}+P_{k-2},\quad k=1,2,\ldots \tag{1, 7}$$

$$Q_{-1}=1,\ Q_0=0,\ Q_k=s_kQ_{k-1}+Q_{k-2}, \tag{1, 8}$$

$$D_k=\begin{vmatrix} P_k & P_{k-1} \\ Q_k & Q_{k-1} \end{vmatrix}; \tag{1, 9}$$

then from (1,7) (1,8) (1,9) it follows that

$$D_k=-D_{k-1},\ \text{and as } D_0=1,\ D_k=(-1)^k. \tag{1, 10}$$

From (1,9) and (1,10) it follows that P_k and Q_k have no other common factors in S than unities and that

$$P_k : Q_k - P_{k-1} : Q_{k-1} = (-1)^k : (Q_kQ_{k-1}). \tag{1, 11}$$

Let a_1 be an arbitrary element $\neq 0$ of K, then we get a uniquely defined sequence of elements $a_1, a_2, \ldots, a_{n+2}$ by the equations

$$a_i=a_i : a_{i+1} \tag{1, 12}$$

From (1,4) we get by multiplying the equations with $a_2, a_3, \ldots$ respectively

$$a_i=s_ia_{i+1}+a_{i+2},\qquad i=1, \ldots, n \tag{1, 4'}$$

From (1,4') (1,7) (1,8) we get

$$\begin{aligned} P_ka_{k+1}+P_{k-1}a_{k+2} &= P_{k-1}(s_ka_{k+1}+a_{k+2})+P_{k-2}a_{k+1} \\ &= P_{k-1}a_k+P_{k-2}a_{k+1}. \end{aligned}$$

By the repeated application of this formula we see that for $i<k$

$$P_ka_{k+1}+P_{k-1}a_{k+2}=P_ia_{i+1}+P_{i-1}a_{i+2}=P_0a_1+P_{-1}a_2=a_1, \tag{1, 13}$$

and by making a similar calculation with the elements Q we get

$$Q_ka_{k+1}+Q_{k-1}a_{k+2}=Q_ia_{i+1}+Q_{i-1}a_{i+2}=Q_0a_1+Q_{-1}a_2=a_2. \tag{1, 13'}$$

Hence

$$\begin{aligned} &(P_ka_{k+1}+P_{k-1}) : (Q_ka_{k+1}+Q_{k-1}) \\ &= (P_ia_{i+1}+P_{i-1}) : (Q_ia_{i+1}+Q_{i-1})=a_1. \end{aligned} \tag{1, 13''}$$

If we multiply (1.13) with Q_{i-1} and (1,13′) with $-P_{i-1}$ and add, then it follows from (1,10) that

$$(-1)^i a_{i+1} = a_1 Q_{i-1} - a_2 P_{i-1}. \tag{1, 14}$$

From (1,12), (1,13′) and (1,10) it follows that

$$a - \frac{P_i}{Q_i} = \frac{(-1)^{i-1}}{a_2} \cdot \frac{a_{i+2}}{Q_i}. \tag{1, 15}$$

Hence if the continued fraction is finite, as $a_{n+2} = 0$

$$a_1 = \frac{P_n}{Q_n}. \tag{1, 15′}$$

If (1, 4) holds, the elements $s_1, s_2, \ldots, s_n$ are said to be the *elements* of the continued fraction ; the quotients $P_i : Q_i$ are the *convergents* and a_{n+1} is called a *complete fraction*. As a_1 is uniquely defined by these elements, we shall denote a_1, if a_{n+1} exists, by

$$a_1 = (s_1, \ldots, s_n \mid a_{n+1}) = \frac{P_n a_{n+1} + P_{n-1}}{Q_n a_{n+1} + Q_{n-1}}, \tag{1, 5′}$$

and if s_n is the last element of the continued fraction,

$$a_1 = (s_1, \ldots, s_n) = \frac{P_n}{Q_n}; \tag{1, 5″}$$

from (1.4), (1,5′), (1,5″) it follows that for $k \leqq n$

$$a_k = (s_k, \ldots, s_n \mid a_{n+1}), \text{ respectively} \tag{1, 5$_k$′}$$
$$a_k = (s_k, \ldots, s_n), \tag{1, 5$_k$″}$$

Let P'_i and Q'_i be defined by

$$\begin{aligned} P'_{-1} &= 0 \quad P'_0 = 1 \quad P'_i = s_{i+i-1} P_{i-1} + P_{i-2} \\ Q'_{-1} &= 1 \quad Q'_0 = 0 \quad Q'_i = s_{i+i-1} Q_{i-1} + Q_{i-2}, \end{aligned} \tag{1, 16}$$

then the following formulae hold :

$$\begin{vmatrix} P'_i & P'_{i-1} \\ Q'_i & Q'_{i-1} \end{vmatrix} = (-1)^i$$

$$\begin{aligned} a_i &= P'_i a_{i+i} + P'_{i-1} a_{i+i+1} \\ a_{i+1} &= Q'_i a_{i+i} + Q'_{i-1} a_{i+i-1} \\ (-1)^i a_{i+i} &= a_i Q'_{i-1} - a_{i+i+1} P'_{i-1} \end{aligned} \tag{1, 17}$$

$$a_i = \frac{P'_i\, a_{i+1} + P'_{i-1}}{Q'_i\, a_{i+1} + Q'_{i-1}}.$$

From

$$\begin{aligned} P_n &= s_n P_{n-1} + P_{n-2} & Q_n &= s_n Q_{n-1} + Q_{n-2} \\ P_{n-1} &= s_{n-1} P_{n-2} + P_{n-3} & Q_{n-1} &= s_{n-1} Q_{n-2} + Q_{n-3} \\ &\cdots\cdots\cdots\cdots & &\cdots\cdots\cdots\cdots \\ P_1 &= s_1 & Q_2 &= s_2 \\ P_0 &= 1 & Q_1 &= 1 \end{aligned}$$

we get the representation of $P_n : P_{n-1}$ and of $Q_n : Q_{n-1}$ as finite continued fractions. Then

$$\frac{P_n}{P_{n-1}} = (s_n, \ldots, s_2, s_1) \quad \frac{Q_n}{Q_{n-1}} = (s_n, \ldots, s_2). \tag{1, 17a}$$

[1/2] There is a very close connection between the finite continued fractions and the algothimus of the *h.c.f.*

Let a_1 be represented by a finite continued fraction. $a_1 = (s_1, \ldots, s_n)$. Then $a_n : a_{n+1} = a_n = s_n$. Hence $a_n = s_n a_{n+1} + 0$, therefore $a_{n+2} = 0$. From (1,13), (1,13'), (1,13''), (1,14) we get therefore

$$\begin{gathered} a_1 = P_n a_{n+1}, \quad a_2 = Q_n a_{n+1} \\ (-1)^n a_{n+1} = a_1 Q_{n-1} - a_2 P_{n-1}. \end{gathered} \tag{1, 18}$$

Hence $a_1 = P_n : Q_n$ belongs to the quotientfield of S.

Every common factor of a_1 and a_2 is a factor of a_{n+1} and a_{n+1} is a common factor of a_1 and a_2. Hence a_1 and a_2 have an *h.c.f.* and this can be represented linearly by a_1 and a_2. Especially $a = P_n : Q_n$ is a representation by two relatively prime elements of S, as

$$P_n Q_{n-1} - Q_n P_{n-1} = (-1)^n.$$

Let every element of the quotientfield of S be representable by a finite continued fraction, and let $s', s'' \neq 0$ be two arbitrary elements of S. then $s' : s'' = a$ can be represented by two relatively prime elements of S so that 1 can be represented in a linear and homogeneous manner by those elements.

Hence $s' : s'' = p : q$ and $pq' + qp' = 1$.

Therefore $s'q = s''p$ and $s''pq' + s''qp' = s'' = q(s'q' + s''p') = qs$.

Hence $s' = ps$. $s'q' + s''p' = s$. So the arbitrary elements s', s'' of S have an *h.c.f.* $(s', s'') = s$ and this is represented in a linear and homogenous manner by s' and s''. From this consideration it follows, that it is not possible to represent the quotients of the elements of an arbitrary *s.r.o.f.* by finite continued fractions, *e.g.*, if in S the elements are factorisable, but the factorisation is not unique there must be a quotient of two elements of S which can be represented by an infinite continued fraction only. As every element of S is represented by a finite continued fraction, the finite continued fractions do not from a field in these cases.

Let a function N(s) which takes positive integral values only, be defined for every element $s \neq 0$ of S and to every pair of elements s, s' of S, let there exist two other elements s_1 and s'' so that

$$s = s_1 s' + s'' \qquad \text{and} \qquad \text{that}$$

$$\text{either} \qquad s'' = 0 \qquad \text{or} \qquad N(s'') < N(s'). \tag{1, 19}$$

Then $s : s' = (s_1 \mid s' : s'') = (s_1, s_2 \mid s'' : s''') = \ldots$, and as

$$N(s') > N(s'') > N(s''') \ldots > 0$$

are all integral numbers, the sequence s', s'', ... must be finite, hence $s : s'$ is a finite continued function. From these considerations we get the following theorem.

Theorem. Let S be an *s.r.o.f.* containing 1, and let a positive integer N(s) be defined satisfying the conditions (1,19) for every element s of S, then the quotientfield of S will be formed by the finite continued fractions of S, and the highest common factor of two elements a_1 and a_2 of S is given by a_{n+1} in (1.18), where P_{n-1}, Q_{n-1} have the significance given by (1,6), (1,7), (1,8).

The formulae (1,13″) and (1,17) are instances of linear fractional substitutions with coefficients from S, the determinant being ± 1. [1/3]

Let A and B be the matrices of substitutions of this kind

$$A = \begin{pmatrix} a_1 & a_2 \\ a_3 & a_4 \end{pmatrix}, \qquad B = \begin{pmatrix} b_1 & b_2 \\ b_3 & b_4 \end{pmatrix} \qquad \begin{matrix} \det A = \epsilon = \pm 1 \\ \det B = \epsilon' = \pm 1 \end{matrix} \tag{1, 20}$$

$$A' = \begin{pmatrix} \epsilon a_4 & -\epsilon a_2 \\ -\epsilon a_3 & \epsilon a_1 \end{pmatrix}, \qquad E = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix};$$

let α be transformed by A into α', and let α' be transformed by B into α'', then it follows

(1) α is transformed by E in α, and $\det E = 1$

(2) α' is ,, ,, A' in α, and $\det A' = \epsilon$

(3) α is ,, ,, AB in α'', and $\det AB = \epsilon\epsilon' = \pm 1$.

The product of two matrices has been defined in Part I, p. 27, and it has been shown in p. 28 that the determinant of a product of matrices is equal to the product of the determinants. The special case which we have to consider here can easily be proved by direct calculation.

An element of K is said to be *equivalent* to α if we get it by transforming α by a linear fractional substitution with determinant ± 1. From (1), (2) and (3) it follows that this equivalence satisfies the conditions of reflexivity symmetry and transitivity (see Part II, [1/3]), and therefore this equivalence defines a partition of K into classes, so that two elements of K are equivalent if and only if they belong to the same class.

By $\begin{pmatrix} s+1 & -1 \\ 1 & 0 \end{pmatrix}$ the element 1 becomes transformed into s, hence all

elements of S are equivalent. From (1,13″) and (1,9) it follows that the elements $a_1, a_2 \ldots$ defined by (1,4) are all equivalent. So more particularly every finite continued fraction $(s_1, \ldots, s_n)$ is equivalent to $a_n = s_n$ and therefore belongs to the class containing the elements of S.

It is sometimes necessary to make the distinction between *proper* and *improper* equivalence. In the first case $\det A = 1$, in the second case $\det A = -1$. As $\det A = \det A'$ in (1, 20) holds the notion of proper equivalence as well as the notion of improper equivalence is a symmetric one. By combining two equivalences of the same kind, we get a proper equivalence, and by combining two equivalences of different kind, we get an improper equivalence. Every element is properly equivalent to itself, for the matrix E has the determinant 1. If in a class of equivalent elements an element α is also improperly equivalent to itself, *i.e.*, if α is transformed into α by E', and $\det E' = -1$, then an arbitrary element β of the same class becomes transformed into α by B and by BE'. One of these matrices has the determinant 1, the other has the determinant -1, so each element of the class is properly and improperly equivalent to α and therefore every element is at the same time properly and improperly equivalent to every other element of the class. If on the other hand α becomes

transformed into β by A as well as by B, where $\det A = 1$, $\det B = -1$, then α becomes transformed into α by BA', where $\det BA' = -1$, and therefore α is improperly equivalent to itself, so that in this case every other element of the class (α) is properly and improperly equivalent to every other element of (α). If in (α) there are no pair of elements properly as well as improperly equivalent, then (α) must be divided in two classes without common elements; the elements of the 1st class are properly, and the elements of the 2nd class are improperly equivalent to α. Elements of the same sub-class are properly equivalent, elements of different sub-classes are improperly equivalent.

Hence

Theorem. In a class of equivalent elements, either every element is properly and improperly equivalent to every other element or there are two sub-classes without common element, so that elements of different sub-classes are improperly equivalent.

As 1 is transformed into itself by $\left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$ every pair of elements is properly and improperly equivalent in the class containing the finite continued fractions.

Let α be transformed into itself by A. Then

$$\alpha = \frac{a_1\alpha + a_2}{a_3\alpha + a_4} \text{ holds, hence } a_3\alpha^2 + (a_4 - a_1)\alpha - a_2 = 0. \qquad (1, 21)$$

There are 3 different cases

1. $a_3 = (a_4 - a_1) = a_2 = 0$. In this case $A = E$ or $-E$.

By these transformations every element is transformed into itself. E and $-E$ generate proper equivalences.

2. (1, 21) is a reducible polynomial in α. This is possible only if α is an element of the quotientfield Q of S.

3. (1, 21) is irreducible. In this case α is algebraic to Q and of order 2 over Q.

From these considerations it follows that not every element is improperly equivalent to itself.

Let the elements $\alpha_1, \alpha_2, \ldots$ as defined by (1, 4) be not all different. *E.g.* let

$$\alpha_i = \alpha_{i+t} \text{ then}$$

$$\alpha_i = (a_{i+1}, \ldots, a_{i+t} \mid \alpha_i) = (a_{i+1}, \ldots, a_{i+t}, a_{i+1}, \ldots, a_{i+t} \mid \alpha_i)$$
$$= \ldots\ldots\ldots\ldots\ldots\ldots$$

Hence α_i can be represented by an infinite periodic continued fraction with the period $s_{i+1}, \dots, s_{i+t}$. From (1, 17) it follows that α_i becomes transformed into itself by the matrix $D = \begin{pmatrix} P'_i & P'_{i-1} \\ Q'_i & Q'_{i-1} \end{pmatrix}$, where $\det D = (-1)^t$, and becomes therefore equivalent to itself by this transformation. From (1, 13′) it follows that $\alpha_1 = \frac{P_{i-1}\alpha_i + P_{i-2}}{Q_{i-1}\alpha_i + Q_{i-2}}$ belongs to the field $Q(\alpha_i)$. Hence α_1 satisfies an equation of degree 2 with co-efficients from S. The same holds for $\alpha_2, \alpha_3 \dots$

[1/4] Let now a periodic sequence $a_1, \dots, a_m, a_1, \dots, a_m, \dots$ of elements of S be given. It is not certain, that in any extension of the quotientfield Q of S there exists an element which can be represented by the infinite periodic continued fraction

$$(a_1, \dots, a_m, a_1, \dots, a_m, \dots). \qquad (1, 22)$$

If such an element exists, it is not certain that this element is uniquely defined in the field. But if there is a field in which there exists one and only one element α represented by the periodic continued fraction (1, 22), then

$$\alpha = (a_1, \dots, a_m, a_1, \dots, a_m, \dots) = (a_1, \dots, a_m \mid \alpha)$$

holds, and this is the case considered just before.

§ 2. Representation of the positive numbers by continued fractions.

[2/1] Let the elements (1, 1) of the set A be the real numbers $\geqq 1$, and let S be the ring of the integers; then the representation by the formulas (1, 3) and (1, 3′)

$$\alpha = s+1 : \alpha' \qquad \text{or} \quad \alpha = s$$

is always possible. If α is not an integral number, there is

$$1 \leqq s < \alpha < s+1, \quad \alpha' = 1 : (s+1-\alpha)$$

and this representation is unique. But if a is an integral positive number,

$a = s'+1$, $s' = 0, 1, 2, \ldots$, there exist two possibilities

$$\begin{aligned} &1. \quad s = s' \,, \; a' = 1 \\ &2. \quad s = s'+1. \end{aligned} \tag{2, 1}$$

Therefore in the representation (1, 4) of any a as a continued fraction

$$\begin{aligned} &s_1 \quad \text{is a non-negative integral number, and} \\ &s_2, s_3, \ldots \text{ are positive integral numbers.} \end{aligned} \tag{2, 2}$$

In order to satisfy (1, 19) we define $N(s) = s$, for $s \geqq 1$. So it follows from (1, 19) that every rational number can be represented by a finite continued fraction. Let a_1 be a rational number $\geqq 1$. If a_1 is an integral number, there are the two representations of a_1 corresponding (2, 1)

$$a_1 = (s'+1)$$

and

$$a_1 = s' + \tfrac{1}{1}.$$

If a_1 is not integral, a_2 is uniquely defined by a_1 ; if a_2 is not integral, a_3 is uniquely defined by a_2, etc. As there is no possibility for alteration in the sequence $a_1, a_2, \ldots$ so long as these elements are not integral, and as there exists a representation by a finite continued fraction, the last complete fraction being integral, there must exist in every representation of a_1 a first integral element a_{r+1} $(r > 0)$. As $a_r = s_r + 1 : a_{r+1}$ is not integral, $1 < a_{r+1} = s'+1$, where $s' > 0$. So we have two possibilities for the continuation of the continued fraction :

$$s_{r+1} = (s'+1) \,, \text{ or } s_{r+1} = s' \,, \; a_{r+1} = 1,$$

and there exist two and only two representations of a_1

$$a_1 = (s_1, \ldots s_r, s'+1), \text{ and } a_1 = (s_1, \ldots s_r, s', 1).$$

If $a_1 > 1$ is irrational, then $s_1 > 0$, and $a_2, a_3 \ldots$ are uniquely defined, none of them being rational ; hence there exists one and only one representation satisfying (2, 2), and this continued fraction is infinite.

If $0 < \beta < 1$, $\dfrac{1}{\beta} = a > 1$, and $\beta = (0 \mid a)$. The essence of these considerations is given by the following theorem :

Theorem. Every positive number can be represented as a continued fraction satisfying the conditions (2, 2). If the number is irrational, the

representation is unique and the continued fraction is infinite. If the number is rational, there exist two representations, one by an even finite continued fraction and the other by an odd finite continued fraction.

[2/2] We will now prove that the converse theorem also holds, *i.e.*, every sequence satisfying (1, 19) defines one and only one real number.

Let $s_1, s_2, \ldots$ be an infinite sequence satisfying the conditions (1, 19). The numbers P and Q should be defined by (1, 7) and (1, 8)

$$P_{-1} = 0,\ P_0 = 1,\ P_1 = s_1,\ P_2 = s_2 s_1 + 1\ ;\ P_k = s_k P_{k-1} + P_{k-2}$$

$$Q_{-1} = 1,\ Q_0 = 0\ Q_1 = 1,\ Q_2 = s_2\ ;\ Q_k = s_k Q_{k-1} + Q_{k-2}.$$

From $P_2 > 0,\ Q_1 > 0$ it follows by mathematical induction that

$$\begin{aligned} &Q_2, Q_3, \ldots, P_3, P_4, \ldots > 0 \quad \text{and that} \\ &0 \leqq P_1 < P_2 < P_3, \ldots < \ldots \\ &Q_0 = 0 < Q_1 \leqq Q_2 < Q_3, \ldots < \ldots \text{ hold.} \end{aligned} \tag{2, 3}$$

The quotients $\frac{P_k}{Q_k}$ are for $k \leqq n$ the convergents of each of the numbers

$(s_1, \ldots, s_n) = \frac{P_n}{Q_n}$. We can therefore apply (1, 15) on $a_1 = \frac{P_n}{Q_n}$.

Hence

$$\frac{P_n}{Q_n} - \frac{P_k}{Q_k} = \frac{(-1)^{k+1}\, a_{k-2}}{a_2 \quad Q_k} \quad \begin{matrix} > 0 & \text{if } k \geqq 1 \text{ odd} \\ < 0 & ,, \ k > 1 \text{ even.} \end{matrix}$$

Hence

$$\frac{P_1}{Q_1} < \frac{P_3}{Q_3} < \ldots < \frac{P_{2m+1}}{Q_{2m+1}}$$

$$\frac{P_2}{Q_2} > \frac{P_4}{Q_4} > \ldots > \frac{P_{2m}}{Q_{2m}} \qquad \text{for } m = 1, 2, 3, \ldots . \tag{2, 4}$$

The quotients $\frac{P_n}{Q_n}$ form therefore two sequences, one is increasing, the other is decreasing, and every number of the first sequence is less than every number of the second one. The intervals $\left(\frac{P_{n-1}}{Q_{n-1}}, \frac{P_n}{Q_n}\right)$ form therefore a

set of intervals ; each of them is included in every other preceding it.

As $$\frac{P_i}{Q_i} - \frac{P_{i-1}}{Q_{i-1}} = \frac{(-1)^i}{Q_i\, Q_{i-1}} \quad . \quad [\text{see } (1, 11)]$$

The length of these intervals converges to 0 . Hence there is for every given sequence (1,19) one and only one real number a so that

$$\frac{P_{2m}}{Q_{2m}} < a < \frac{P_{2m-1}}{Q_{2m-1}} \qquad \text{holds for } m = 1, 2, \ldots ; \qquad (2, 5)$$

a is a positive number ; this is the number which becomes *defined* by the sequence $s_1, s_2 \ldots$. a can be represented by a continued fraction. We will prove that this continued fraction is equal to $(s_1, s_2 \ldots)$. This is not obvious ; from the above it only follows that if there exist a number whose representation as a continued fraction is $(s_1, s_2, \ldots)$, then this number can only be the number a defined by that sequence, as there exist no other number situated within all the intervals $\left(\frac{P_n}{Q_n}, \frac{P_{n+1}}{Q_{n+1}}\right)$, but it may so happen that there is no such number and that the representation of a as a continued fraction, furnishes another sequence which defines the same real number a. Of course this complication cannot occur. We will show it by the following lemma, which gives us some idea of the distribution of the continued fractions on the axis of the real numbers.

Lemma. Let $P_i : Q_i$ be the convergents of $(s_1, s_2, \ldots)$, and $P' : Q'$ be the last convergent of $(s_1, \ldots, s_{n-1}, \ldots, s_n + t)$, where $n = 1, \ldots$ and $t > 0$, then

$$P_n : Q_n < P_{n-1} : Q_{n-1} \leqq P' : Q' \qquad \text{if } n \text{ is odd,}$$
$$\text{and} \qquad P_n : Q_n > P_{n-1} : Q_{n-1} \geqq P' : Q' \qquad \text{if } n \text{ is even.} \qquad (2, 6)$$

Proof.

$$Q_{n+1} = s_{n+1}\, Q_n + Q_{n-1}$$

$$Q' = (s_n + t)\, Q_{n-1} + Q_{n-2} = Q_n + t\, Q_{n-1}$$

$$\frac{1}{t}\, Q' = \frac{1}{t}\, Q_n + Q_{n-1} \leqq Q_{n+1}.$$

The equality holds only if $n = 0$, or if $t = s_{n+1} = 1$.

Applying (1, 11) we get

$$\frac{P_{n+1}}{Q_{n+1}} - \frac{P_n}{Q_n} = \frac{(-1)^{n+1}}{Q_n\, Q_{n+1}} = \frac{(-1)^{n+1}\, t}{Q_n\,(Q'+a)}, \text{ where } a \geqq 0 \qquad (2, 7)$$

$$\frac{P'}{Q'} - \frac{P_n}{Q_n} = \left(\frac{P'}{Q'} - \frac{P_{n-1}}{Q_{n-1}}\right) - \left(\frac{P_n}{Q_n} - \frac{P_{n-1}}{Q_{n-1}}\right)$$

$$= \frac{(-1)^n}{Q'Q_{n-1}} - \frac{(-1)^{n-1}}{Q_n\, Q_{n-1}} = (-1)^{n-1}\, \frac{Q'-Q_n}{Q'Q_n\, Q_{n-1}} = \frac{(-1)^{n+1}\, t}{Q_nQ'}, \qquad (2, 7')$$

From (2, 7) and (2, 7') the proposition (2, 6) follows directly.

A real number, defined by any continued fraction $C = (s_1, \ldots, s_n, s_{n+1}, \ldots)$ belongs to the interval $(P_n : Q_n\ ,\ P_{n+1} : Q_{n+1})$ as an interior point, and the numbers defined by $C' = (s_1, \ldots, s_{n-1}, s_n + t,\ s'_{n+1}, \ldots)$ are interior points of $(P_{n-1} : Q_{n-1}\ ,\ P' : Q')$ From the preceding theorem it follows that these intervals have no common interior point; they have a common endpoint only if $1 = t = s_{n+1}$. Of course $(s_1, \ldots, s_n, 1) = (s_1, \ldots, s_n + 1)$. However if one of the continued fractions C and C' is infinite, C and C' must define different numbers. As every pair of different continued fractions can be written in the form C, C' it follows that a real number which is defined by an infinite continued fraction cannot be defined by another one, and must therefore be an irrational number. Hence an irrational number can be defined only by its representation (1, 5). But as every infinite sequence satisfying (1, 19) defines a real number, we get the following theorem.

Theorem. Every sequence $s_1, s_2, \ldots$ satisfying (1, 19) is furnished by the representation of a suitable number a, and different real numbers correspond to different sequences. The value of a becomes defined by the inequalities (2, 5).

[2/3] From the lemma we see how the continued fractions are distributed on the axis of the real numbers. For abbreviation we will write for an arbitrary finite set $s, \ldots, s_n$

$$(s_1, \ldots, s_n + t) = [n \mid t] \qquad t = 0, 1, 2 \ldots$$

$$(s_1, \ldots, s_n, u) = [n, u] \qquad u = 1, 2, \ldots .$$

Let n be odd

$$[n \mid 0] < [n, u] < \ldots < [n, 2] < [n, 1] = [n \mid 1] < [n \mid 2] < \ldots$$

If n is even, the notation $<$ must be replaced by $>$. For the continued fractions having only one element s_1, there is $(s_1) = s_1$. So we get a first partition of the real axis by the numbers 0, 1, 2, 3 ... . The continued fractions $(s_1 \mid \omega)$ beginning with s_1 are situated between s_1 and $s_1 + 1$.

E.g. the continued fractions $(3, s_2)$ are $3 + 1 = 4, 3 + \frac{1}{2}, 3 + \frac{1}{3}$ ... these having 3 as limiting point. The continued fractions $(3, 2, s_3)$ are situated between $(3, 2) = 3 + \frac{1}{2}$ and $(3, 3) = 3 + \frac{1}{3}$; $(3, 2, 1) = (3, 3) < (3, 3, 2), < \ldots$ the limiting point is the upper limit. If we continue this procedure, the intervals become subdivided, every segment (a, b) of the nth division is divided by the $(n + 1)$th division into sub-segments ; there is in every segment only one limiting point and this is a if n is odd and b if n is even.

$P_1 = 3 \quad P_2 = 7 \quad P_3 = 7s_3 + 3 \qquad P_3 : Q_3 = 3\frac{1}{2} - \frac{1}{2} : (2s_3 + 1)$

$Q_1 = 1 \quad Q_2 = 2 \quad Q_3 = 2s_3 + 1$

$s_3 =$	1	2	3	4
$\frac{1}{2} : (2s_3 + 1) =$	$\frac{1}{3}$	$\frac{1}{10}$	$\frac{1}{14}$	$\frac{1}{18}$

As every rational number is represented by a finite continued fraction, and especially by a continued fraction with an even as well as with an odd number of elements, it will occur as a point of subdivision and it will be the limiting point of one odd subdivision and of one even subdivision. So we get a natural classification of the rational positive numbers: Numbers being representable by $P_1 : Q_1$, those being representable by $P_2 : Q_2$, etc.

The importance of this classification becomes clear by the following theorem.

Theorem. If $s > 0$, and $\dfrac{P_{2n-1}}{Q_{2n-1}} < \dfrac{r}{s} < \dfrac{P_{2n}}{Q_{2n}}$ then

$s > Q_{2n} > Q_{2n-1}$.

Proof. From the supposition it follows directly that

$$0 < \frac{r}{s} - \frac{P_{2n-1}}{Q_{2n-1}} < \frac{P_{2n}}{Q_{2n}} - \frac{P_{2n-1}}{Q_{2n-1}} = \frac{1}{Q_{2n} Q_{2n-1}}$$

and as s and Q_{2n-1} are positive

$$0 < r\,Q_{2n-1} - s\,P_{2n-1} < \frac{s}{Q_{2n}}.$$

the middle part of this inequality is an integral positive number,

Hence $1 < \dfrac{s}{Q_{2n}}$ *i.e.* $Q_{2n} < s$.

because 1 is the least of all positive integers.

This theorem means that the best approximations of a real number by the help of rational numbers with limited positive denominators are the approximations by the convergents $P_n : Q_n$.

§ 3. Periodic continued fractions with integral co-efficients.

3/1] In [1/4] it has been shown that if a periodic continued fraction in S represents one and only one element α of K, this element α sati-fies a quadratic equation in S. In the special case where S is the ring of the integers, it has been shown that the sequences (2, 2) represent one and only one (positive) real number ; so, it follows that the periodic continued fractions (2, 2) represent elements quadratic to the field R of the rational numbers. However the converse also holds.

Theorem. If a positive irrational number α belongs to a field Λ, where $[\Lambda : R] = 2$ then α is represented by a periodic continued fraction.

Proof. α has to be the root of polynomial

$$ax^2 + 2bx + c \; ; \qquad (3, 1)$$

on the other hand α can be represented by a continued fraction $\alpha = (s_1, \ldots, s_n, \lambda)$. From (1, 13'') it follows that

$$\alpha = \frac{P_n\lambda + P_{n-1}}{Q_n\lambda + Q_{n-1}} \; ;$$

hence $a(P_n\lambda + P_{n-1})^2 + 2b(P_n\lambda + P_{n-1})(Q_n\lambda + Q_{n-1}) + c(Q_n\lambda + Q_{n-1})^2 = 0$

or λ is a root of $\qquad gx^2 + 2hx + k = 0, \qquad (3, 2)$

where $$\begin{vmatrix} g & h \\ h & k \end{vmatrix} = \begin{vmatrix} P_n & P_{n-1} \\ Q_n & Q_{n-1} \end{vmatrix}^2 \begin{vmatrix} a & b \\ b & c \end{vmatrix} = \begin{vmatrix} a & b \\ b & c \end{vmatrix} . \qquad (3, 3)$$

Let β be the second root of (3, 1) and μ be defined by

$$\mu = \frac{-Q_{n-1}\beta + P_{n-1}}{Q_n \;\; \beta - P_n}, \qquad \text{hence} \qquad (3, 4)$$

$$\beta = \frac{P_n\mu + P_{n-1}}{Q_n\mu + Q_{n-1}},$$

As $a\beta^2 + 2b\beta + c = 0$, the number μ satisfies (3, 2). From (3, 4) it follows that

$$\mu = \frac{-Q_{n-1}}{Q_n} \pm \frac{1}{Q_n(Q_n\beta - P_n)} .$$

As $$\left| a - \frac{P_n}{Q_n} \right| < \left| \frac{P_{n+1}}{Q_{n+1}} - \frac{P_n}{Q_n} \right| = \frac{1}{Q_n Q_{n+1}} < \frac{1}{Q_n^2},$$

$$a = \frac{P_n}{Q_n} + \frac{\epsilon}{Q_n^2}, \quad \text{where } |\epsilon| < 1.$$

So $$P_n = Q_n a - \frac{\epsilon}{Q_n}.$$

From (3, 2) it follows therefore

$$\mu = \frac{-Q_{n-1}}{Q_n} \pm \frac{1}{Q_n^2(\beta - a) + \epsilon}$$

$$= \frac{-Q_{n-1}}{Q_n} \left(1 \pm \frac{1}{Q_n Q_{n-1}(\beta - a) + \epsilon} \right) \qquad (3, 5)$$

$$< 0 \quad \text{for } Q_n Q_{n-1} |\beta - a| > 2.$$

Hence after a certain integer n, the root μ becomes negative, therefore $k : g = \lambda\mu$ becomes negative. As $b^2 - a c = h^2 - kg$ admits only a finite number of solutions h, k, g for which $-kg > 0$, there must be different complete fractions satisfying the same equation (3, 2) with $-kg > 0$. These equations have only one positive root. Therefore the same complete fraction will be repeated. Hence the continued fraction is periodic.

The period of a continued fraction will be denoted by a bar ; *e.g.* [3/2]

$$(s_1, \ldots, s_m, \overline{s_{m+1}, \ldots, s_{m+\ell}}) \qquad (3, 6)$$

denotes a continued fraction with the period $s_{m+1}, \ldots, s_{m+\ell}$. If $m = 0$, the continued fraction is said to be *purely periodic*.

Let $a_1, a_2, \ldots$ be the complete fractions of (3, 6). As $a_i = a_{i+\ell}$, for $i > m$, every property which holds for complete fractions of sufficient high index, holds for every complete fraction of index $> m$. From (3, 4) we know that the root conjugate to a complete fraction of sufficiently high index is negative. Therefore this property holds for every complete fraction of index $> m$, *i.e.*, for every purely periodic continued fraction. In a purely periodic continued fraction $s_1 \neq 0$, hence the continued fraction represents a number > 1.

A root λ of a quadratic equation is said to be *reduced* if $\lambda > 1$, and the conjugate root satisfies $0 > \mu > -1$. We will prove now that every purely periodic continued fraction represents a reduced quadratic number.

Let

$$\alpha = \overline{(s_1, \dots, s_n)} = (s_1, \dots, s_n \mid \alpha) \qquad (3, 7)$$

$$\xi = \overline{(s_n, \dots, s_1)} = (s_n, \dots, s_1 \mid \xi) \qquad (3, 7')$$

be two purely continued fractions, the elements s_i being the same in both continued fractions, but ordered in an inverse manner.

Let $P_i : Q_i$ be the convergents of α. Then (see 1, 17[a])

$$\frac{P_n}{P_{n-1}} = (s_n, \dots, s_1), \quad \frac{Q_n}{Q_{n-1}} = (s_n, \dots, s_2)$$

hence $\xi = \dfrac{P_n \xi + Q_n}{P_{n-1} \xi + Q_{n-1}}$ holds.

Let $\beta = \dfrac{-1}{\xi}$, then $0 > \beta > -1$, and

$P_{n-1} \beta^{-2} + (P_n - Q_{n-1}) \beta^{-1} - Q_n = 0$ hence β is a root of

$f(x) = Q_n x^2 + (Q_{n-1} - P_n) x - P_{n-1}$.

As $\alpha = \dfrac{P_n \alpha + P_{n-1}}{Q_n \alpha + Q_{n-1}}$, α is also a root $f(x)$, and as $\alpha > 1$, the roots α and β are different. The essence of these considerations is therefore:

Theorem. If α is represented by a purely continued fraction (3.7), it is a reduced quadratic number; let β be the number conjugate to α, and $\xi = -\beta^{-1}$, then ξ is represented by (3, 7').

Every continued fraction is equivalent to its complete fractions, especially every periodic continued fraction is equivalent to a purely continued fraction; hence

Corollary. Every quadratic number is equivalent to a reduced quadratic number.

[3/3] In order to find out the representation of any quadratic number by a continued fraction, we represent this number

$$\alpha = \frac{a + \sqrt{D}}{b} = s + \frac{1}{\alpha}, \text{ where } s < \alpha < s + 1,$$

and $a, b, s, D > 0$ are integral numbers.

$$\alpha' = -\frac{a' + \sqrt{D}}{b'}, \quad \text{where } a' = bs - a, \quad b' = (D - a'^2) : b.$$

From these formulas we can find out by a simple numerical scheme the numbers a,b; a',b'...defining uniquely the complete fractions $\alpha, \alpha',\ldots$ and the numbers $s, s', \ldots$ defining the continued fraction. As this continued fraction is periodic, one pair a, b must be repeated after a finite number of steps. Then the first period is finished and the calculation has to be stopped.

Examples.

1. $\alpha = \dfrac{-1+\sqrt{5}}{2}$ (harmonic section) $D=5$

a	b		s
-1	2	$0 < \dfrac{-1+\sqrt{5}}{2} < 1$	0
1	$\dfrac{5-1^2}{2}=2$	$1 < \dfrac{1+\sqrt{5}}{2} < 2$	1
1	2		

The last complete fraction is therefore equal to the preceding ; hence $\alpha = (0, \overline{1})$. This is the simplest continued fraction, but the worst for practical calculation, *viz.* the numbers P_s, Q_s are increasing slower than in any other case,

2. $\alpha = \sqrt{26}$.

a	b	s
0	1	5
5	1	10
5	1	

hence $\alpha = (5,\overline{10})$. This example is very convenient for quick and exact calculation.

$P_0 =$	1	$Q_0 =$	0
$P_1 =$	5	$Q_1 =$	1
$P_2 =$	51	$Q_2 =$	10
$P_3 =$	515	$Q_3 =$	101
$P_4 =$	5201	$Q_4 =$	1020
$P_5 =$	52525	$Q_5 =$	10301
$P_6 =$	530451	$Q_6 =$	104030

Hence $a = \frac{530451}{104030} - \epsilon$, where $0 < \epsilon < 10^{-11}$. Therefore $a = 5{,}099019513(60)$ the last two figures being uncertain.

As the error $\epsilon = \left| a - \frac{P_n}{Q_n} \right| < \frac{1}{Q_n Q_{n-1}} < \frac{1}{s_{n+1} Q_n^2}$, it is useful to stop the calculation just before a large s_{n+1}.

3. $a = \sqrt{2}$ $D = 2$

a	b	s
0	1	1
1	1	2
1	1	

Hence $2 = (1, \overline{2})$. As P_4, Q_4 are increasing very slowly we will use another method.

$\sqrt{2} = \sqrt{200} : 10$. If $\sqrt{200} = \frac{P_n}{Q_n} \pm \epsilon$, $\sqrt{2} = \frac{P_n}{10Q_n} \pm \frac{\epsilon}{10}$

So we represent $\sqrt{200}$ by a continued fraction. $D = 200 = 14^2 + 4$.

a	b		s
0	1	$14 < \sqrt{200} < 15$	14
14	4	$7 < \frac{14 + \sqrt{200}}{4} < 8$	7
14	1	$28 < 14 + \sqrt{200} < 29$	28
14	4		

Hence $\sqrt{200} = (\overline{14, 7, 28})$,

$P_0 = 1$	$Q_0 = 0$
$P_1 = 14$	$Q_1 = 1$
$P_2 = 99$	$Q_2 = 7$
$P_3 = 2786$	$Q_3 = 197$
$P_4 = 19601$	$Q_4 = 1386$

$\sqrt{200} = \frac{19601}{1386} - \epsilon$, $0 < \epsilon < \frac{1}{Q_4 Q_5} < \frac{1}{28 \cdot Q_4} 2 < 3 \cdot 10^{-8}$

$\sqrt{2} = \frac{1960{,}1}{1386} - \epsilon'$, $0 < \epsilon' < 3 \cdot 10^{-9} = 1{,}41421356(4)$,

true to eight figures after the decimal point.

Exercises. Prove that $(a, \overline{2a}) = \sqrt{a^2+1}$, and calculate $\sqrt{2501}$, $\sqrt{82}$, $\frac{\sqrt{7}+2}{3}$, $\sqrt{17}$. Calculate $\sqrt{3}$ directly and also by help of $\sqrt{300}$.

In order to prove the converse of the last theorem, it is useful to consider the following lemma.

Lemma. If $\alpha > 1$ and $\beta < 0$ are conjugate quadratic numbers, $\alpha = (s, s_1, s_2, \ldots)$, then all the complete fractions $\alpha_1 = (s_1, s_2 \ldots)$, $\alpha_2 = (s_2, \ldots)$ are reduced numbers.

Proof. $\alpha = s + \frac{1}{\alpha_1}$, $\beta = s + \frac{1}{\beta_1}$, α_1 and β_1 are conjugate numbers $\alpha_1 > 0$, $\frac{-1}{\beta_1} = s - \beta > s \geqq 1$ hence α_1 is reduced, and by repetition of this procedure it follows that $\alpha_2, \ldots$ are reduced.

Theorem. Every reduced quadratic number is represented by a purely periodic continued fraction.

Proof. Every quadratic number is represented by a periodic continued fraction $(a, \ldots, s, \overline{s_1, \ldots, s_n})$. Let this number be reduced and let the periodicity of the continued fraction begin with s_1 only (*i.e.* let $s \neq s_n$), then it follows from the last lemma that $(\overline{s, s_1, \ldots, s_n})$ is a reduced number too. We will prove that this is impossible. Using the same notations as in the lemma we state

$$\alpha_1 = \alpha_{n+1}, \qquad \text{hence } \beta_1 = \beta_{n+1}$$

$$\alpha = s + \frac{1}{\alpha_1}, \quad \alpha_n = s_n + \frac{1}{\alpha_{n+1}} \qquad \text{hence } \beta = s + \frac{1}{\beta_1}, \quad \beta_n = s_n + \frac{1}{\beta_{n+1}}$$

$$\frac{-1}{\beta_1} = s - \beta, \quad \frac{-1}{\beta_{n+1}} = s_n - \beta_n ;$$

but as α and α_{n-1} are reduced, $0 < -\beta < 1$, and $0 < -\beta_{n-1} < 1$ hold; hence $s - 1 < \frac{1}{\beta_1} < s$, and $s_n - 1 < \frac{-1}{\beta_{n+1}} < s_n$. From $\beta_1 = \beta_{n+1}$ it follows therefore that $s = s_n$.

Theorem. Let $\alpha = \sqrt{\frac{r}{t}} > 1$ be irrational, then [3/5]

$$\alpha = (s, \overline{s_1, \ldots, s_n}), \qquad (3, 8)$$

and $$s_n = 2s$$
$$s_i = s_{n-i} \qquad i = 1, \ldots, n-1 \qquad (3, 8')$$

hold. If conversely (3, 8) and (3, 8') hold, then α is an irrational square-root > 1.

Proof. As $\alpha > 1$ and the number $\beta = -\alpha < 0$ is conjugate to α, it follows from the lemma that the complete fractions $\alpha_1, \alpha_2, \ldots$ of α are reduced, and therefore periodic. Hence α satisfies (3, 8). As $\alpha = s + \frac{1}{\alpha_1}$, and $-\alpha = \beta = s + \frac{1}{\beta_1}$, α_1 and β_1 are conjugate.

$$\alpha_1 = \overline{(s_1, \ldots, s_n)}\,. \qquad (3, 9)$$

Therefore it follows from the theorem of [3/2] that

$$-1 : \beta_1 = \overline{(s_n, \ldots, s_1)} \qquad (3, 9')$$

$0 = \alpha_1 + \beta_1 = (s + \alpha_1) + 1 : \beta_1$. Hence $-1 : \beta_1 = s + \alpha_1$.

Therefore $$\overline{(s_n, \ldots, s_1)} = (2s, \overline{s_1, \ldots, s_n}) \qquad (3, 10)$$

holds. (3, 8') is equivalent to (3, 10). Conversely if α_1 is defined by (3, 8), (3, 8'), then (3, 10) holds.

Let α_1 and β_1 be defined by (36) and (36'), and let $\beta = s + 1 : \beta_1$; then it follows from (3, 9) and (3, 9') that α_1 and β_1 and therefore α, β are conjugate, and from (3, 10) it follows that $\alpha + \beta = 0$. Hence α and β are the roots of a rational polynomial $tx^2 + 0\,x - r$. From $0 \neq s_n = 2s$, it follows that $s \geqq 1$, and therefore $\alpha > 1$, and as (3, 8) is an infinite continued fraction, α must be irrational.

Corollary. Let $\alpha = \sqrt{r} : t$ and $P_1 : Q_1, \ldots$ be the convergents of α, then

$$t\,P_{kn}^2 - r\,Q_{kn}^2 = (-1)^{kn}\,t \qquad (3, 11)$$

holds for every $k = 1, 2, \ldots$

Proof. Let $\alpha_1, \ldots$ be the complete fractions of α, then $\alpha_i = \alpha_{i+sn}$

$$\alpha_{kn} = s_{kn} + \frac{1}{\alpha_{kn-1}} = 2s + \frac{1}{\alpha_1} = s + \alpha.$$

But, as $a = \dfrac{P_{2n}a_{2n} + P_{2n-1}}{Q_{2n}a_{2n} + Q_{2n-1}}$, (see 13′)

$$a = \frac{P_{2n}(s + a) + P_{2n-1}}{Q_{2n}(s + a) + Q_{2n-1}}$$ holds ; hence

$$Q_{2n}\,a^2 - P_{2n}\,s - P_{2n-1} + a\,(Q_{2n}\,s + Q_{2n-1} - P_{2n}) = 0.$$

As $a^2 = r : t$ is rational, and a is irrational

$$P_{2n-1} + P_{2n}\,s - Q_{2n}\,r : t = 0$$

$$-P_{2n} + Q_{2n}\,s + Q_{2n-1} = 0$$

hold. If we multiply these equations by $Q_{2n}t$, respectively $-P_{2n}t$ and add, we get $tP_{2n}^2 - r\,Q_{2n}^2 + t(P_{2n-1}\,Q_{2n} - Q_{2n-1}\,P_{2n}) = 0$ and from this formula we get (3, 11) directly.

§ 4. Applications on Theory of Numbers.

It is proposed to solve [4/1]

$$ax - by = 1 \qquad (4, 1)$$

by integral x and y.

Obviously (4,1) cannot be solved if there is a common factor of a and b different from ± 1. We therefore suppose a and b to be relatively prime. $a : b$ can be represented by an even continued fraction (see [2/i]).

$a : b = (s_1, \ldots, s_{2m})$

$a : b = P_{2m} : Q_{2m}$, and as a and b are positive and relatively prime

$a = P_{2m}$ $b = Q_{2m}$ and therefore

$$a\,Q_{2m-1} - b\,P_{2m-1} = P_{2m}Q_{2m-1} - Q_{2m}P_{2m-1} = (-1)^{2m} = 1$$

holds. Hence we get the integral solutions by

$$x = Q_{2m-1} + k\,b,$$
$$y = P_{2m-1} + k\,a \qquad \text{where } k = 0, \pm 1, \pm 2, \ldots$$

To solve $x^2 - dy^2 = 1$ (*Pell's equation*) by integral x and y, we will use [4/2] (3, 11).

$$\sqrt{d} = a = (s, \overline{s_1, \ldots, s_n})\ ;$$

then $P_{2n}^2 - d\,Q_{2n}^2 = (-1)^{2n}$

Therefore if n is even, $(x, y)=(P_{kn}, Q_{kn})$

and if n is odd, $(x, y)=(P_{2kn}, Q_{2kn})$

are solutions for every positive integral k.

E.g. $x^2-26\,y^2=1.$

$$\sqrt{26}=(5, \overline{10}) \qquad n=1.$$

By this method we get the solutions

$$\begin{aligned}(x, y)&=(P_2, Q_2)=(51,\ 10)\\&=(P_4, Q_4)=(5201,\ 1020)\\&=(P_6, Q_6)=(530451,\ 104030)\\&\ldots\ldots\ldots\ldots\ldots\ldots ,\end{aligned}$$

§ 5. CONTINUED FRACTIONS WHOSE ELEMENTS ARE $\phi(x)$.

[5/1] In § 3 and § 4 the system of the positive numbers has been taken for the system A, S being the ring of the integers. We will now consider another system A.

Let K be an arbitrary field and x an indefinite not included in K. The elements of K will be denoted by a, b, c, d, with and without indices. The elements of the ring $K[x]$ will be denoted by

$$f(x),\ g(x), \tag{5, 1}$$

The ring (sub-ring of a field) $K[x]$ will be set as the ring S.

In order to get a new system A we create new elements denoted by Greek letters

$$\phi(x),\ \psi(x),\ \chi(x),\ \omega(x) \tag{5, 2}$$

in the following manner.

$$\phi(x)=a_n x^n+a_{n-1}x^{n-1}+\ldots+a_0+a_{-1}\,x^{-1}+\ldots+a_{-k}\,x^{-k}+\ldots+$$

$$=0\,x^{n+m}+\ldots+0\,x^{n+1}+a_n x^n+\ldots+a_{-k}x^{-k}+\ldots=\sum_{k=n}^{-\infty} a_k x^k. \tag{5, 3}$$

This is a purely formal definition. It means that to every sequence of co-efficients from K with fixed decreasing integral indices

$$a_n,\ a_{n-1}\ \ldots.$$

there corresponds one of our new elements, and this element will not be changed if we take before a_n a finite set of null-coefficients. Nothing has been supposed about convergence. We have to define the addition and the multiplication of the elements (5, 2), and we will define them in such a way that the elements (5, 2) for which $a_k = 0$ for $k < 0$ form a subring isomorphic to K $[x]$.

So we define :

$$\text{Let } n \geqq m,\ \phi(x) = \sum_{k=n}^{-\infty} a_k x^k$$

$$\psi(x) = \sum_{k=m}^{-\infty} b_k x^k = \sum_{k=n}^{-\infty} b_k x^k, \text{ where } 0 = b_{m+1} = \ldots b_n \text{ if } n > m.$$

$$\text{then } \phi(x) + \psi(x) = \chi(x) = \sum_{k=n}^{-\infty} c_k x^k$$

$$\phi(x).\ \psi(x) = \omega(x) = \sum_{k=n+m}^{-\infty} d_k x^k$$

$$\text{where} \qquad c_k = a_k + b_k, \qquad d_k = \Sigma a_i\ b_{k-i} \qquad (5, 4)$$

$$n \geqq i \geqq k - m.$$

The definitions (5, 4) are obviously independent of null-coefficients put before ; the commutative, associative and distributive laws hold, and the subtraction is uniquely defined by

$$b_k = c_k - a_k$$

Hence for the null-element every co-efficient is 0. If for the co-efficients of $\psi(x)$ the conditions $b_k = 0$, for $k \neq 0$ hold, then in $\phi(x)\,\psi(x) = \Sigma\, d_k\, x^k$

$$d_k = b_0\, a_k \qquad \text{holds.}$$

The elements (5, 2) form a ring R and those elements for which $a_{k<0} = 0$ form a subring for which the addition and multiplication has been defined in the same way as for polynomials. Hence there is an isomorphism I by which this subring becomes isomorphic to K$[x]$. Let $\phi(x) \neq 0$, then $\phi(x)$ has at least one co-efficient $\neq 0$, let n be the highest index of the non-vanishing co-efficients, then

$$\phi(x) = \sum_{n}^{-\infty} a_k x^k a_n \neq 0.$$

n is said to be the degree of $\phi(x)$. From (42) it follows directly:

The degree of a product is equal to the sum of the degrees of the factors.

The degree of a sum of elements of different degree is equal to the maximum degree of the summands.

From this remark it follows that a product of two elements $\neq 0$ cannot be equal to 0. Hence the ring of the elements (5,2) is an *s.r.o.f.* We will now *identify* the elements of K $[x]$ with the corresponding elements (5, 2).

So the elements $\sum_{k=0}^{-\infty} b_k x^k$, $b_k = 0$, for $k \neq 0$ become identified with b_0,

and $b_0 \phi(x) = \sum_{k}^{-\infty} b_0 a_k x^k$ holds.

Let $\psi_k(x) = x^{-k} + 0\, x^{-k-1} + 0 \ldots$; then $x^k \psi_k(x) = 1$. Every field containing R contains the quotient field Q of K $[x]$. The elements of R, which are quotients of elements of K (x) form a ring which is isomorphic to a subring of Q. The elements of this ring will therefore be identified with the corresponding elements of Q. So $\psi_k(x)$ becomes identified with x^{-k} and the *finite* sum $\sum_{k=m}^{-m} a_k x^k$ becomes identified with the *symbolic* sum $\phi(x) = \sum_{k=m}^{-\infty} a_k x^k$, where $a_k = 0$, for $k < -m$.

Using these notations we can extend the algorithmus of division of the polynomials to the elements (5, 2).

Let $\phi(x)$ and $\psi(x)$ be of degree n and m respectively and a_n and b_m their highest co-efficients,

$\frac{a_n}{b_m} = c_{n-m}$. $\phi_1(x) = \phi(x) - c_{n-m} x^{n-m} \psi(x)$ is of degree $n_1 < n$.

By repetition of this procedure we get

$$\phi_2(x) = \phi_1(x) - c_{n_1-m} x^{n_1-m} \psi(x) = \phi(x) - (c_{n-m} x^{n-m} + c_{n_1-m} x^{n_1-m})\, \psi(x),$$

and by further repetitions we get an enumerable set of elements c_k of K

$k \leqq n-m$, so that $-\chi_\nu(x) = \sum_{k=n-m}^{n_\nu - m} c_k x^k$, and

$\phi_{\nu+1}(x) = \phi(x) - \chi_\nu(x)\, \psi(x)$. $\chi_\nu(x)$ is of degree $n_{\nu+1} < n_\nu$.

Let $\chi(x) = \sum_{n=m}^{-\infty} c_n x^n = \chi_\nu(x) + \omega_\nu(x)$;

then $\omega_\nu(x)$ is of degree $n_{\nu+1}-m$ and therefore $\psi(x)\omega_\nu(x)$ is of degree $n_{\nu+1}$.

$\phi(x)-\psi(x)\chi(x)=\phi(x)-\psi(x)\chi_\nu(x)-\psi(x)\omega_\nu(x)$ is of degree $<n_\nu$ for every ν, hence this difference is 0. Hence $\phi(x)=\psi(x)\chi(x)$ holds. As $\psi(x)$ was supposed to be an arbitrary element $\neq 0$ of R, it follows:

Theorem. The set R of the elements (5, 2) is a field containing the quotientfield Q of K [x].

$$\text{Let } \phi(x)=\sum_{n}^{-\infty} a_s x^s=\sum_{n}^{0} a_s x^s+\sum_{-1}^{-\infty} a_s x^s=f(x)+\phi_1(x) \qquad [5/2]$$

This representation of $\phi(x)$ as a sum of polynomial in x and an element which is zero or of degree <0 is obviously unique.

As $1:\phi_1(x)$ is of degree >0 if and only if $\phi_1(x)$ is of degree <0 it follows that there is one and only one representation $\phi(x)=f(x)$ or $\phi(x)=f(x)+1:\psi(x)$, where the degree of $\psi(x)$ is >0

So we can apply [1/1] [1/2], [1/3] to the case when the elements

$$a, a', a'', \ldots, a_1, a_2, \ldots$$

of A are the elements (5, 2) of degree >0, and $s, s', s'', \ldots, s_1, s_2, \ldots$ are the polynomials in x. [See (1, 1), (1, 2)]

If we make these suppositions about A and S, then $s_1, s_2, \ldots, a_2, \ldots$ are uniquely defined by (1, 4), and so we get a unique representation of the elements (1, 1) as a continued fraction, the elements s_i being polynomials in x of degree >0. If we permit s_1 to be an element of K (*i. e.*, a polynomial of degree 0 or the null element) then we get a unique representation of every element (5, 2) as a continued fraction. Hence

Theorem. The elements (5, 2) can be represented in one and only one manner by a continued fraction $(s_1, s_2 \ldots)$ where s_i is a polynomial in x, whose degree is >0, for $i>1$.

The degree of a polynomial s has just the same properties as the function $N(s)$ in [1/2]. From that section and the preceding theorem we get therefore the

Corollary. The finite continued fractions represent the elements of Q and every element of Q is represented by a finite continued fraction.

We will now use the representation of real numbers by continued fractions in order to *approximate* the elements (5, 2).

[5/8] As (1, 15) holds,

$$a_1 - \frac{P_k}{Q_k} = \frac{(-1)^{k-1}}{a_2} \quad \frac{a_{k+2}}{Q_k};$$

we will prove that the right side of this equation is an element, whose degree decreases indefinitely as k increases.

The elements a_i have been defined by $a_i = a_i : a_{i+1}$, a common factor being arbitrary (see (1, 12)) hence $a_2 : a_{k+2} = a_2 \cdot \ldots \cdot a_{k+1}$; therefore

$$\text{degree}\left(a_1 - \frac{P_k}{Q_k}\right) = -\text{degree } Q_k - d_2 - \ldots - d_{k+1}. \qquad (5, 5)$$

$d_k > 0$ for $k > 1$ and $a_k = s_k + 1 : a_{k+1}$. (see 1, 4). As the degree of a sum of summands of different degrees is equal to the highest of the degrees of the summands degree $(s_k) = \text{degree } (a_k) = d_k$ for $k > 1$.

$$Q_1 = 1,\ Q_2 = s_2,\ \ldots,\ Q_i = s_i\ Q_{i-1} + Q_{i-2}.$$

Hence degree $(Q_k) > 1$ for $k > 1$, hence the degrees increase with the index and therefore

$$\text{degree } Q_i = \text{degree } (s_i Q_{i-1}) = d_i + \text{degree } (Q_{i-1}) = \sum_{k=2}^{i} d_k \qquad (5, 6)$$

Hence from (5, 5) and (5, 6) it follows that

$$\text{degree}\left(a_1 - \frac{P_k}{Q_k}\right) = d = -2 \sum_{j=2}^{k} d_j - d_{k+1} =$$

$$\text{degree}\left(\frac{-1}{Q_k Q_{k+1}}\right) \leqq 1 - 2k. \qquad (5, 7)$$

The efficiency of the approximation of an irrational number a by a continued fraction became clear by the theorem that if $\frac{a}{b}$ approximates a better than $\frac{P_k}{Q_k}$, then $b > Q_k$ holds. The corresponding theorem holds in the case we consider here.

Theorem. If $f(x)$ and $g(x)$ are polynomials of K $[x]$,

$$\beta = \frac{f(x)}{g(x)} - \frac{P_k(x)}{Q_k(x)} \neq 0, \quad \text{and}$$

$$d' = \text{degree}\left(a_1 - \frac{f(x)}{g(x)}\right) < d = \text{degree}\left(a_1 - \frac{P_k}{Q_k}\right), \text{ then}$$

$$\text{degree } g(x) > \text{degree } Q(x) \text{ holds.} \qquad (5, 8)$$

Proof. $\beta = \left(a_1 - \frac{P_k}{Q_k}\right) + \left(\frac{f(x)}{g(x)} - a_1\right)$. The two summands on the right side have different degrees hence $d' = \text{degree } \beta$ holds. As $\beta\, Q_k\, g(x)$ is a polynomial, $0 \leqq \text{degree } (\beta\, Q_k\, g(x)) = \text{degree } g(x) + \text{degree } (Q_k) + d'$

$$\text{degree } g(x) \geqq -\text{degree } Q_k - d' > -\text{degree } Q_k - d = \text{degree } Q_k - d_{k+1}$$

$$> \text{degree } Q_k \text{ [see (5, 7), (5, 8)].}$$

This theorem enables us to approximate functions given by a power series of $\frac{1}{x}$ by rational function in the neighbourhood of $x = 1 : 0$.

Exercise. $\frac{1}{2} \log \frac{x+1}{x-1} = x^{-1} \frac{1}{3} x^{-3} + \frac{1}{5} x^{-5} + \ldots$

Represent this function by a continued fraction and approximate it by rational functions.

Lemma. Let $a = (s_1, \ldots)$, $a' = (s'_1, \ldots)$; let m be the lowest index for which $s_m \neq s'_m$ holds and $(s_1, \ldots, s_m) = A$, $(s'_1, \ldots, s'_m) = A'$ then [5/4]

$$\text{degree } (a - a') = \text{degree } (A - A') \quad \text{holds.} \qquad (5, 9)$$

Proof. Without loss of generality we suppose that degree $s_m = r \geqq$ degree $s'_m = r'$. We shall use the ordinary notations for the convergents of a and for those of a' we shall use them with a dash.

$P_i = P'_i$

$Q_i = Q'_i$ for $i < m$

$Q_m = s_m Q_{m-1} + Q_{m-2}$ degree $Q_m = r + q$

$Q'_m = s'_m Q_{m-1} + Q_{m-2}$ degree $Q'_m = r' + q$

$$A - A' = \left(\frac{P_{m-1}}{Q_{m-1}} + \frac{(-1)^m}{Q_{m-1}\,Q_m}\right) - \left(\frac{P_{m-1}}{Q_{m-1}} + \frac{(-1)^m}{Q_{m-1}\,Q'_m}\right)$$

$$= (-1)^m \frac{s_m - s'_m}{Q_m\,Q'_m}.$$

1. If $r = r'$, degree $(s_m - s'_m) \geqq 0$

degree $(A - A') \geqq -2(d + q) = \text{degree } \frac{1}{Q'^2_m}$;

2. If $r > r'$, degree $(s_m - s'_m) = r$,

degree $(A - A') \geqq r - (r + q + r' + q) \geqq 2(r' + q) = \text{degree } \frac{1}{Q'^2_m}$. Hence

$$\text{degree } (A - A') \geqq \text{degree } \frac{1}{Q'^2_m} \qquad (5, 10)$$

holds in every case.

From (5, 9) it follows that

degree $(A - a) = \text{degree } \frac{1}{Q_m\,Q_{m+1}} < \text{degree } \frac{1}{Q_m^2} \leqq \text{degree } \frac{1}{Q'^2_m}$,

and that degree $(A' - a') = \text{degree } \frac{1}{Q'_m\,Q'_{m+1}} < \text{degree } \frac{1}{Q'^2_m}$. Hence

degree $(a - a') = \text{degree } [(A - A') - (A - a) - (A' - a')] = \text{degree } (A - A')$

viz. the degree of the first one of the three summands is greater as the degrees of the two other summands.

Theorem. Let $s_1, s_2, \ldots$ be an infinite set of polynomials of $K[x]$ and let for $i > 1$, degree $s_i > 0$, then there exists a continued fraction $(s_1, s_2, \ldots)$.

Proof. The set $s_1\ s_2, \ldots$ defines uniquely the values

$P_1, \ldots, Q_1, \ldots$, and $P_N : Q_N = (s_1, \ldots, s_N)$.

Let $1 < n < N$.

From the preceding lemma it follows, that degree $(P_N : Q_N - P_n : Q_n)$

$$= \text{degree } (P_{n+1} : Q_{n+1} - P_n : Q_n) = \text{degree } \left(\frac{1}{Q_n\,Q_{n+1}}\right) = -k_n,$$

where k_n increases to infinity, with n.

$$P_N : Q_N = \sum_{k=m}^{-\infty} a_k x^k = \sum_{k=m}^{1-k_n} b_k x^k + \sum_{k=-k_n}^{-\infty} c_k x^k \qquad (5, 11)$$

The co-efficients b_k are independent of N. As k_n increases with the index n, we get an infinite set $b_m, \ldots, b_{-k_1}, \ldots$ defining

$$\phi(x) = \sum_{\nu=m}^{-\infty} b_\nu x^\nu$$

Finally we have to prove that $\phi(x) = (s_1, s_2, \ldots)$. Let $\phi(x) = (s'_1, s'_2, \ldots)$, and m be the smallest index for which $s_m \neq s'_m$, then it follows from the lemma that for every $n > m$,

degree $(\phi(x) - (s_1, \ldots, s_n))$ = degree $(\phi(x) - (s_1, \ldots, s_m))$ holds.

But $\phi(x) - (s_1, \ldots, s_n) = b_{k_n} x^{-k_n} + b_{k_n+1} x^{-k_n-1} + \ldots$ is of degree $-k_n$ and decreases infinitely with n.

§ 6. Continued Fractions with Rational Elements.

Let S be the field of the rational numbers, then every finite continued fraction [6/1]

$$(s_1, \ldots, s_n) = \frac{P_n}{Q_n} = \frac{P_1}{Q_1} + \left(\frac{P_2}{Q_2} - \frac{P_1}{Q_1}\right) + \ldots + \left(\frac{P_n}{Q_n} - \frac{P_{n-1}}{Q_{n-1}}\right) \qquad (6, 1)$$

$$= s_1 + \frac{1}{Q_1 Q_2} + \ldots + \frac{(-1)^n}{Q_{n-1} Q_n}$$

represents a rational number, but an infinite continued fraction defines a number if and only if

$$\Sigma \frac{(-1)^n}{Q_{n-1} Q_n} \qquad (6, 2)$$

converges. If the sum (6, 2) is convergent,

$$(s_1, s_2, \ldots) \qquad (6, 3)$$

defines a real number equal to (6, 2). A necessary condition for the convergence of (6, 3) is therefore

$$| Q_{n-1} Q_n | \longrightarrow \infty \qquad (6, 4)$$

If the numbers Q_n are either > 0 each, or < 0 each, or of alternating sign, the sum (6, 2) becomes an alternating sum ; hence the continued fraction converges if $| Q_n Q_{n-1} |$ increases steadily and (6, 4) is satisfied.

[6/2] *Theorem.* If $\Sigma | s_n |$ is convergent, (6, 3) is divergent.*

Proof. We prove by mathematical induction that

$$Q_n < \prod_{j=1}^{n} (1 + | s_j |) \qquad (6, 5)$$

As $Q_1 = 1$, $Q_2 = s_2$, the formula holds for $n < 3$. If (6, 5) is true for $n < m$,

$$Q_m = s_m Q_{m-1} + Q_{m-2}$$

$$| Q_m | \leq \prod_{j=1}^{m-2} (1 + | s_j |) \cdot \{ | s_m | (1 + | s_{m-1} |) + 1 \}$$

$$\leq \prod_{j=1}^{m} (1 + | s_j |).$$

If Σs_j converges, the infinite product $\Pi (1 + | s_j |)$ converges to a positive number Q, and $| Q_n | < Q$ holds for every index n. Hence (6, 4) does not hold and the continued fraction is divergent.

[6/3] Let $s_i > 0$ for $i > 1$. From $Q_1 = 1$, $Q_2 = s_2 > 0$, $Q_n = s_n Q_{n-1} + Q_{n-2}$ it follows by mathematical induction that each number $Q_i > 0$.

$$Q_n Q_{n-1} = s_n Q^2_{n-1} + Q_{n-1} Q_{n-2} > Q_{n-1} Q_{n-2} > 0.$$

(6, 2) is therefore an alternating series, whose elements have steadily decreasing absolute values. This series converges therefore if and only if (6, 4) is satisfied. These considerations lead to the following

Theorem. Let $s_i > 0$ for $i > 1$, then the continued fraction (6, 3) is convergent if and only if Σs_i is divergent.

Proof. If $\Sigma s_i = \Sigma | s_i |$ is convergent, the continued fraction is divergent, as it has been proved by the preceding theorem.

* We use the term " divergent " for every non-convergent series.

Let Σs_i be divergent, then $s_i \to \infty$. As $Q_i > 0$, $Q_1 = 1$, and $Q_{2n+1} = s_{2n+1} Q_{2n} + Q_{2n-1}$, we get by mathematical induction that $Q_{2n+1} \geqq 1$. Hence $Q_{2n} = s_{2n} Q_{2n-1} + Q_{2n-2} \geqq s_{2n} + Q_{2n-2}$, and as $Q_2 = s_2$, it follows by mathematical induction that

$$Q_{2n} \geqq \sum_{2}^{2n} s_{2j}.$$

Hence

$$Q_{2n-1} Q_{2n} \to \infty, \quad \text{and} \quad Q_{2n} Q_{2n+1} \to \infty,$$

therefore (6, 4) is satisfied, and as we stated above, this condition is sufficient for the convergence of (6, 3) in the case considered here.

Let $s_1 = 0$, $s_i \geqq 1$ for $i > 1$, then $s_i \to \infty$, and it follows from the preceding theorem that the continued fraction (6, 3) converges to a value in the interval [6/4]

$$\left(\frac{P_1}{Q_1}, \frac{P_2}{Q_2} \right) = (0, 1)$$

We will show that this value is *irrational*.

Let $\alpha_1 = (s_1, s_2, \ldots)$ be rational, say $\alpha_1 = \frac{a_2}{a_1}$, where a_1, a_2 are integral.

then $a_1 > a_2$ and $\frac{a_2}{a_1} = \frac{1}{s_2 + \alpha_2}$ hence $\alpha_2 = \frac{a_1}{a_2} - s_2 = \frac{a_3}{a_2}$, where $a_3 = a_1 - s_2 a_2$ is integral. as $\alpha_2 = (0, s_3, \ldots)$, there is $0 < \alpha_2 < 1$; hence $a_2 > a_3 > 0$. In the same manner, we get

$$\alpha_2 = \frac{1}{s_3 + \alpha_3}, \quad \alpha_3 = (0, s_4, \ldots) = \frac{a_4}{a_3}, \text{ and } a_3 > a_4 > 0.$$

By repetition of this procedure we get an infinite set of decreasing integral positive numbers

$$a_1 > a_2 > a_3 > a_4 > \ldots.$$

and that is impossible.

Hence α_1 is irrational.

Example. $s_1=0,\ s_2=2,\ s_3=1,$ and for $m>1$

$$s_{2m}=\frac{2\,.\,4\ldots 2m-2}{1\,.\,3\ldots 2m-3}>1,\quad s_{2m+1}=\frac{3\,.\,5\ldots 2m-1}{2\,.\,4\ldots 2m-2}>1$$

then it is

$$Q_1=1,\ Q_2=2,\ Q_3=3,$$

and from the identities

$$2\,.\,4\ldots 2m=1\,.\,3\ldots(2m-1)\,.\,s_{2m}+2\,.\,4\ldots 2m-2$$

$$1\,.\,3\ldots 2m+1=2\,.\,4\ldots 2m\,.\,s_{2m+1}+1\,.\,3\ldots 2m-1,$$

it follows by mathematical induction that

$$Q_{2m}=2\,.\,4\ldots 2m,\quad Q_{2m+1}=1\,.\,3\ldots 2m+1$$

hence

$$Q_n\,.\,Q_{n-1}=n\,!$$

the continued fraction is irrational. Its value is

$$\frac{1}{Q_1\,Q_2}-\frac{1}{Q_2\,Q_3}+\ldots+\frac{(-1)^n}{Q_{n-1}\,Q_n}+\ldots=\Sigma(-1)^n:n\,!=e^{-1}.$$

Hence e is irrational.

PART IV.

APPROXIMATE SOLUTION

§ 1. Horner's scheme.

Let K be an arbitrary field, *e.g.*, the field of the real numbers or [1/1] of the complex numbers, q be an element of K and $f(x)$ a polynomial of $K[x]$.

$$f(x)=\sum_{i=0}^{n} a_i x^i=(x-q)\sum_{i=1}^{n} a'_i\, x^{i-1}+a'_0=(x-q)f_1(x)+a'_0$$

then $a_i=a'_i-q\ a'_{i-1}$. Hence

$$\begin{aligned} a'_n &= a_n \\ a'_{n-1} &= a_{n-1}+qa'_n \qquad (1,1)\\ &\cdots\cdots\cdots\cdots \\ a'_0 &= a_0+qa'_1 \end{aligned}$$

holds. We can arrange the calculation of the coefficients a'_i as follows:

$$\begin{array}{ccccccc} a_n & a_{n-1} & a_{n-2} & \cdots & a_1 & a_0 & \\ & qa'_n & qa'_{n-1} & \cdots & qa'_2 & qa'_1 & \qquad (1,1')\\ \hline a'_n & a'_{n-1} & a'_{n-2} & & a'_1 & a'_0 & \end{array}$$

$$f(q)=a'_0 \qquad (1,2)$$

We can find out by the same method $f_{11}(x)$ satisfying

$$f_1(x)=(x-q)\ f_{11}(x)+a''_1\ .\ f_{11}(x)=\sum_{i=2}^{n} a''_i x^{i-2}\ .$$

After $n-1$ steps we get $f(x)$ represented as a polynomial in $x-q$

$$f(x)=a'_0+a''_1(x-q)+\ldots+a^{(n)}_{n-1}(x-q)^{n-1}+a_n(x-q)^n.$$

This representation is known in Analysis as the *Taylor series* of $f(x)$ at the point $x=q$. The successive calculation of the coefficients can easily be done on using the last line of (1, 1') up to a'_1. The complete scheme for this calculation is called *Horner's scheme*. It is the most convenient method

for calculating $f(q)$ if q and the coefficients of $f(x)$ are given, and furnishes the representation of $f(x)$ by $(x-q)$. The calculation will be explained by the following

Example. $f(x)=x^4-15x^3+68x^2-119x+67$.

$q=1$,

1	−15	68	−119	67
	1	−14	54	−65
1	−14	54	−65	2
	1	−13	41	
1	−13	41	−24	
	1	−12		
1	−12	29		
	1			
1	−11			

$$\begin{aligned} f(x) &= (x^3-14x^2+54x-65)(x-1)+2 \\ &= (x^2-13x+41)(x-1)^2-24(x-1)+2 \\ &= (x-12)(x-1)^3+29\ (x-1)^2-24(x-1)+2 \\ &= (x-1)^4-11\ (x-1)^3+29\ (x-1)^2-24(x-1)+2. \end{aligned}$$

[1/2] Horner's scheme is very useful for calculating the roots. The method will be explained by the help of the above example.

$x=y+1$, $f(x)=g(y)=y^4-11y^3+29y^2-24y+2$

$f(1)=g(0)=2$

$f'(1)=g'(0)=-24$

We therefore suspect that there is a root of $f(x)$ near $x=1$. In the neighbourhood * of $x=1$ $f(x)\backsim -24y+2$; $f(x)=0$ for $y\backsim 0{\cdot}1$. Therefore we represent $g\ (y)$ by a polynomial in $y-0{\cdot}1$.

$q=0{\cdot}1$

1	−11	29	−24	2
	+ 0·1	− 1·09	+ 2·791	−2·1209
1	−10·9	27·91	−21·209	−0·1209
	+ 0·1	− 1·08	+ 2·683	
1	−10·8	26·83	−18·526	
	+ 0·1	− 1·07		
1	−10·7	25·76		
	+ 0·1			
1	−10·6			

* The sign $\backsim$ means "approximately equal to."—We apply here an elementary theorem on continuous functions, which will be cited in [2/2].

As $f(1{\cdot}1) = -0{\cdot}1209 < 0$, there is a root between 1 and 1·1. We approximate therefore $f(x)$ by

$$-18{\cdot}526\ (x-1{\cdot}1) - 0{\cdot}1209, \qquad \text{hence} \quad x - 1{\cdot}1 \backsimeq -0{\cdot}007.$$

$q = -0{\cdot}007$	1	−10·6	25·76	−18·526	−0·1209
		− 0·007	+0·074249	− 0·180839743	+0·150947878201
	1	−10·607	25·834249	−18·706839743	0·010047878201
		− 0·007	+0·074298	− 0·181359829	
	1	−10·614	25·908547	−18·888199572	
		− 0·007	+0·074347		
	1	−10·621	25·982894		
		− 0·007			
	1	−10·628			

Hence the root is approximately equal to 1·093. The next approximation is $q = 0{\cdot}00053$. If we would continue in exactly the same manner, the calculation would become very burdensome. We therefore neglect the terms q^4 and q^3, and we get a good approximation by taking terms up to q^2. To this approximation we are led by the following consideration. $18{\cdot}888199572q = 0{\cdot}010047878201 + 25{\cdot}982894q^2 - 10{\cdot}628q^3 + q^4$. As $q \backsimeq 5{\cdot}3.10^{-4}$, the two last terms will influence only the 9th and the following decimals of the right side for $5{\cdot}3\ 10^{-4} < q < 5{\cdot}4.10^{-4}$ the quadratic term becomes 0·000007...

Hence $q = 0{\cdot}0005324$. $x = 1{\cdot}0935324$.

On using this approximation we could easily get some more figures of this decimal development. The most difficult task is sometimes to get a first approximation of the roots. For this purpose it is often helpful to know the values of $f(x)$ for a suitable set of values x. We may get these values by Horner's scheme, but it is often useful to abbreviate this scheme in following manner.

Given $q_1, q_2, \ldots, q_m < s$. We calculate by Horner's scheme [1/3]

$$f(x) = b_1 + (x-q_1)\ f_1(x) \qquad f(q_1) = b_1$$
$$f_1(x) = b_2 + (x-q_2)\ f_2(x) \qquad f(q_2) = b_1 + (q_2-q_1)\ b_2$$
$$\ldots \qquad \ldots \qquad \ldots \qquad \ldots \qquad \ldots$$
$$f_{m-1}(x) = b_m + (x-q_m)\ f_m(x)$$
$$f(x) = b_1 + b_2(x-q_1) + b_3(x-q_1)(x-q_2) + \ldots + b_m(x-q_1)\ldots(x-q_{m-1}) + (x-q_1)\ldots(x-q_m)\ f_m(x).$$

We will develop the polynomial considered in the previous example in this manner.

$q_1=1$	1	−15	68	−119	67
		1	−14	54	−65
$q_2=2$	1	−14	54	−65	2
		2	−24	60	
$q_3=3$	1	−12	30	−5	
		3	−27		
	1	−9	3		

$$f(x)=2-5(x-1)+3(x-1)(x-2)+(x-1)(x-2)(x-3)(x-9)$$

$$f(0)=67,\ f(1)=2,\quad f(2)=-3,\quad f(3)=-2,\quad f(9)=2+8(-5+3.7)=130.$$

This representation shows that for $x<1$, $f(x)>0$, *viz.* each of the terms becomes >0, and that for $x>9$, $f(x)>0$, *viz.* the 1st, the 4th and the sum of the two other terms become >0. So all roots are situated in the interval (1, 9). We calculated one root in the interval (1, 2); there is at least one root in the interval (3, 9). $f(8)=2-5\cdot7+7\cdot6\,(3-5)=-117$. Hence there is a root in the interval (8, 9). The readers may calculate it by Horner's scheme as an *exercise*.

[1/4] We will use here a different method of calculation. If $f(x)=\sum_0^n a_s x^s=0$, $\frac{1}{x}$ satisfies the condition $\sum_0^n a_{n-s}\,(\frac{1}{x})^s=0$, and to every root of x in the interval (0, 1) there corresponds a value of $\frac{1}{x}>0$. These considerations lead to the following method of approximation due to *Lagrange*.

If ξ is a root of $g(x)$, $a<\xi<a+1, =a+\frac{1}{1}$, $g(x)=f(a-x)=g_1(\frac{1}{a-x})$, $\eta_1>1$ is a root of g_1, and $b\leqq\eta_1<b+1$, $\eta_1=b+\frac{1}{\eta_2}$. By repetition of this procedure we will get a representation of ξ as a continued fraction. If we stop the calculation after n steps we get the approximation $\frac{P_n}{Q_n}$ and the error becomes

$$\xi-\frac{P_n}{Q_n}\quad<\quad\frac{1}{\ldots_n Q_{n+1}}\quad<\quad\frac{1}{Q_n^2}\ .$$

This method will be illustrated by the example previously used. We know that $x^4-15x^3+68x^2-119x+67$ has a root in the interval (8, 9). Therefore we represent this polynomial by Horner's method as a polynomial in $x-8$.

$q=8$	1	−15	68	−119	67
		8	−56	96	−184
	1	− 7	12	−23	−117
		8	8	160	
	1	1	20	137	
		8	72		
	1	9	92		
		8			
	1	17			

Hence $117\,\eta_1^4-137\,\eta_1^3-92\,\eta_1^2-17\,\eta_1-1=0$.

By mental arithmetic we see that for $q=2$ the last coefficient becomes positive, but for $q=1$ it becomes negative, so η is in the interval (1, 2), and we have to make the Horner-development for $q=1$.

$q=1$	117	−137	−92	−17	−1
		117	−20	−112	−129
	117	−20	−112	−129	−130
		117	97	− 15	
	117	97	−15	−144	
		117	214		
	117	214	199		
		117			
	117	331			

In the same manner as it has been done for η_1, we see that η_2 is situated in the interval (1, 2).

$q=1$	130	144	−199	−331	−117
		130	274	75	−256
	130	274	75	−256	−373
		130	404	479	
	130	404	479	223	
		130	534		
	130	534	1013		
		130			
	130	664			

η_3 is in the interval (2, 3).

$q=2$					
	373	−223	−1013	−664	−130
		746	1046	66	−1196
	373	523	33	−598	−1326
		746	2538	5142	
	373	1269	2571	4544	
		746	4030		
	373	2015	6601		
		746			
	373	2761			

Probably the reader has by now the experience that q has to be chosen so that the sign of the last coefficient does not change, but that the procedure adopted for $q=1$ would alter this sign; $q=4$ will alter the sign of the second coefficient in the second main row of Horner's scheme, but the third coefficient will not change its sign, 6601 being too big. Hence the following coefficients will increase and therefore will not become negative. However for $q=5$, -6601 is counterbalanced by more than 3000, and therefore -2761 by more than 12000, and so the sign of the last coefficient would be altered. Hence $q=4$.

$q=1$					
	1326	−4544	−6601	−2761	−373
		5304	3040	−14244	−68020
	1326	760	−3561	−17005	−68393
		5304	24256	82780	
	1326	6064	20695	65775	
		5304	45472		
	1326	11368	66167		
		5304			
	1326	16672.			

The next q will become $=1$. Hence $\xi=(8, 1, 1, 2, 4, 1, \ldots)$.

P_1	= 8	Q_1	= 1
P_2	= 9	Q_2	= 1
P_3	= 17	Q_3	= 2
P_4	= 43	Q_4	= 5
P_5	= 189	Q_5	= 22
P_6	= 232	Q_6	= 27
		Q_7	$\geqq$ 49

$\xi \sim \frac{232}{27}$, the error $\frac{232}{27} - \xi$ is positive and $\leqq \frac{1}{27 \cdot 49} = 0{\cdot}00075 \ldots$

As $\frac{232}{27} = 8{\cdot}5925 \ldots$, the value of ξ is true up to the second decimal only ; the third decimal may be 2 or 1. From this example the reader will see that Lagrange's method is sometimes not very convenient for practical calculation.

In the last sub-sections Horner's scheme has been used for the solution of equations with real coefficients by real roots. But the scheme can be applied—as it has been stated on the beginning of this section—for arbitrary fields. We will use it now to find out a theorem on complex numbers. Let $b_0,\ b_1+b_0,\ \ldots,\ b_n + \ldots + b_0$, be the coefficients of a polynomial, [1/5]

$$\begin{array}{lllll} q=a & b_n & b_1+b_n & b_2+b_1+b_n & \ldots \quad b_n+b_{n-1}+\ldots+b_n \\ & & ab_n & ab_1+(a+a^2)\,b_n & \ldots \quad ab_{n-1}+\ldots \Sigma a^i\, b_n \\ \hline b_n & b_1+ & b_2+(1+a)b_1+ & \ldots & \frac{b_n(1-a)+\ldots+b_n(1-a^{n+1})}{1-a} \\ & (1+a)\,b_n & (1+a+a^2)\,b_n & & \end{array}$$

Hence if a is a root, $\sum_0^n b_i = \sum_0^n b_i\, a^{n+1-i}$ (1, 3)

Let b_i be positive numbers and a complex.

1.) If $|a| < 1$, the equation (3) cannot hold.

2.) If $|a| = 1$, $a = \cos v + i \sin v$,

$\Sigma b_i = \Sigma b_i \cos (n+1-k)v$, hence $\cos (n+1-k)v = 1$, $v = 0$, $a = 1$ but in this case the last coefficient is equal to $\Sigma (n+1-k)\ b_i > 0$; hence $|a| > 1$. If therefore in $a_0 y^n + a_1 y^{n-1} + \ldots + a_n$

$$a_0 < a_1 < \ldots < a_n , \qquad (1, 4)$$

then for every root a of this polynomial $|a| > 1$ holds. Hence if $y = \frac{1}{x}$, the roots β of $\Sigma a_i\, x^i$ satisfy $|\beta| < 1$. This theorem is known as

Kakeya's theorem. The complex roots of $\Sigma a_i\, x^i$ have all absolute values < 1, if the coefficients satisfy (1, 4).

§ 2. The roots of real polynomials.

[2/1] In this section

$$a,\quad b,\quad c,\quad d,\quad e \tag{2, 1}$$

—with or without indices and dashes—denote real numbers; in the same manner

$$\alpha,\quad \beta,\quad \gamma,\quad \delta$$

$$\bar{\alpha},\quad \bar{\beta},\quad \bar{\gamma},\quad \bar{\delta} \tag{2, 2}$$

denote complex numbers, α and $\bar{\alpha}$ etc. being conjugate.

Hence $\alpha+\bar{\alpha}$ is real; $\alpha\bar{\alpha}$ is positive; $\alpha-\bar{\alpha}=ci$, and if $c=0$, α is real.

$$f(x)=a_0+a_1x+\ \dots\ +a_{n-1}x^{n-1}+x^n \tag{2, 3}$$

can be represented by $f(x)=\prod_{i=1}^{n}(x-\alpha_i)$ (2, 4)

(see Part II [13/2]).

Theorem. If α is a root of (2, 4), $\bar{\alpha}$ is also a root of (2, 4).

1st Proof. Let K be the field of real numbers, i and $-i$ be the roots of x^2+1, then $K(i)=K(-i)$ is the field of the complex numbers and there is an automorphism J of this field interchanging i with $-i$ and leaving the real numbers unaltered. $f(x)$ will not be altered by J, hence α will be transformed into a root of $f(x)$, but as α will be transformed to $\bar{\alpha}$, the theorem is true.

2nd Proof. If α is real, $\bar{\alpha}=\alpha$. If α is not real, $(x-\alpha)(x-\bar{\alpha})=g(x)$ is a real polynomial and irreducible in the field of the real numbers. As $f(x)$ and $g(x)$ have a common root, these polynomials have a common factor of positive degree. Hence $f(x)$ is divisible by $g(x)$ and $\bar{\alpha}$ is therefore a root of $f(x)$.

Corollary 1.

$$f(x)=(x-c_1)\ \dots\ (x-c_r)(x-\beta_1)(x-\bar{\beta}_1)\ \dots\ (x-\beta_k)(x-\bar{\beta}_k) \tag{2, 5}$$

where $n=r+2k$.

Corollary 2. If every root of $f(x)$ is counted as many times as its order of multiplicity in (2, 4), the number of the real roots is $\equiv n \pmod 2$.

Corollary 3. If n is odd, there exists at least one real root.

$$\Delta = \prod_{i<j} (a_i - a_j)^2 = F(a_1, \ldots, a_n) \tag{2, 6}$$

is called the *discriminant* of $f(x)$. As F is a symmetric polynomial in $a_1, \ldots, a_n$ with integral coefficients, it follows (see Part II [10/3]) that

$$\Delta = g(a_1, \ldots, a_n) \tag{2, 7}$$

where g is a polynomial with integral coefficients. From (2, 6) it follows (see Part II, [10/3]) that

$$\Delta = 0 \quad \text{if and only if} \quad a_i = a_j \text{ for } i \neq j. \tag{2, 8}$$

Let the n roots a_j be all different. As it has been proved in Part II [10/4],

$$\Delta = \delta^2, \quad \delta = \begin{vmatrix} a_1^{n-1} & \ldots & a_1 & 1 \\ \ldots\ldots & \ldots & \ldots & \ldots \\ a_n^{n-1} & \ldots & a_n & 1 \end{vmatrix}. \tag{2, 9}$$

To get $\bar{\delta}$ we have to interchange every number with its conjugate. From (2, 5) it follows that this operation means k interchanges of rows in the determinant (2, 9).

Hence $\bar{\delta} = (-1)^k \delta$, and therefore

$$(-1)^k \Delta = (-1)^k \delta^2 = \delta\bar{\delta} > 0.$$

Hence the following theorem holds.

Theorem. Let $f(x)$ of degree n have n different roots. Then the discriminant of $f(x)$ is positive (negative) when the number of pairs of conjugate non-real roots is even (odd).

Corollaries. A real polynomial of degree 3 has three real roots if and only if the discriminant is positive.

A real polynomial of degree 4 with positive discriminant has either four different real roots or two pairs of conjugate complex roots.

Exercise. Prove the preceding theorem without the help of (2, 8).

[2/2] If we consider x as a real variable, $f(x)$ is a real continuous and differentiable function. Hence:

1. If $a < b$, and the signs of $f(a)$ and $f(b)$ are different, then there is a root of $f(x)$ in the interval (a, b).

2. If $a < b$, and $f(a)=f(b)=0$, then there is a root of $f'(x)$ in the interval (a, b).

3. If $f(x)=(x-a)^k g(x)$, $g(a) \neq 0$, $k > 0$, then

$$f'(x)=(x-a)^{k-1} g_1(x), \quad g_1(a) \neq 0.$$

For, $g_1(x)=k\, g(x)+(x-a)\, g'(x)$.

Therefore if the roots of $f(x)$ take m different real values $a_1, \ldots, a_m$, there exists at least one root of $f'(x)$ in each of the $m-1$ different intervals (a_i, a_{i+1}). If a_i is a multiple root with the multiplicity $q+1$, it can be considered as a set of q degenerated intervals, each of them containing exactly one root of $f'(x)$. $f(x)$ has the same sign at every point of an interval; in two consecutive intervals (a_{i-1}, a_i) and (a_i, a_{i+1}) the sign of $f(x)$ is different, when a_i is a simple root or a multiple root of an odd order; the sign is not different if a_i is a multiple root of even order. Thus if a_i is a multiple root of order $q+1$ of $f(x)$, it is a multiple root of order q of $f'(x)$.

These properties hold for every analytic function with a finite number of roots, and are not special properties of polynomials. If the coefficients of $f(x)$ are all positive (all negative), $f(x)$ becomes obviously positive (negative) for every positive value of x. Hence $f(x)$ is not zero for any positive value of x. Thus

$f(x)$ has no positive roots, if there is no change of the signs in the sequence of the coefficients of $f(x)$. So we are led to study the connection between the existence of roots and the signs of the coefficients. The experience we got in using Horner's scheme will be very helpful to us.

By developing $f(x)$ as a polynomial in $x-b$, we get

$$f(x)=f_b(x-b)=a_{b,n}\,(x-b)^n+\ \ldots\ +a_{b,0}\,. \tag{2, 10}$$

So to every real number b and the fixed function $f(x)$ there belongs a set of $n+1$ real numbers $a_{b,n}, a_{b,n-1}, \ldots, a_{b,0}$

such that $$a_{b,n}=a_n \quad \text{independent of } b \tag{2, 11}$$

$$a_{b,0}=f(b).$$

Hence $a_{b,n} = 0$ if and only if b is a real root of $f(x)$, and for $0 < k < n$

$$a_{b,k} = \frac{1}{k!} f^{k}(b). \tag{2, 12}$$

The sequence

$$a_{b,n}, \quad a_{b,n-1}, \ \ldots \ a_{b,0} \tag{2, 13}$$

may be reduced by striking out every element equal to 0. If in this reduced sequence $a_{b,n}$ has a sign different from the sign of the preceding element, it contributes one *change*. We will consider the number $C(b)$ of these changes.

$$0 \leqq C(b) \leqq n \tag{2, 14}$$

If the first and the last element of the reduced sequence have the same sign, $C(b)$ is an even number ; if they have different signs, $C(b)$ is odd. From (2, 11) it follows therefore that $C(b)$ is even in the intervals where $a_n f(b) > 0$ and it is odd where $a_n f(b) < 0$.

If we strike out in (2, 13) an element which contributes no change but is different from the first, the number of changes will be unaltered ; if we strike out an arbitrary element, the number of changes will not increase. In order to investigate the function $C(b)$ we will consult Horner's scheme. We will compare the changes in the rows

$$a_n, \quad a_{n-1}, \ \ldots \ a_1, \ a_0$$
$$a'_n, \ a'_{n-1}, \ \ldots, \ a'_1, \ a'_0$$

where $\qquad a'_n = a_n \quad a'_k = a_k + q a'_{k-1}.$

Let $q > 0$. If a'_k contributes a change in the 2^d row, then $a'_k \neq 0$ and it is of the same sign as a_k. Hence the 2^d row has the same number of changes as the sub-sequence formed by those elements a_k of the first row, which have the same sign as the corresponding elements of the second row. The number of changes of the 2^d row is therefore less or equal to the number of the changes in the first. The same holds for every pair of subsequent rows in the following system which we get by Horner's scheme.

$$\begin{array}{lllll}
a_n, & a_{n-1}, & \ldots, & a_2, & a_1, \quad a_0 \\
a'_n, & a'_{n-1}, & \ldots, & a'_2, & a'_1, \quad a'_0 \\
a''_n, & a''_{n-1}, & \ldots, & a''_2, & a''_1, \quad a'_0 \\
a'''_n, & a'''_{n-1}, & \ldots, & a'''_2, & a''_1, \quad a'_0 \\
\ldots\ldots & \ldots\ldots & \ldots\ldots & \ldots\ldots & \ldots\ldots \\
a^{(n)}_n, & a^{(n)}_{n-1}, & \ldots, & a'''_2, & a''_1, \quad a'_0
\end{array}$$

But the last row is identical with

$$a_{q,n},\ a_{q,n-1},\ \ldots,\ a_{q,0}.$$

Hence for $q > 0$, $C(q) \leqq C(0)$ holds. (2, 15)

Let $c = b + q > b$

$$f(x) = f_b(x-b) = f_c([x-b]-q),$$

then it follows from (2,15) that the number of changes in f_c is not greater than the number of changes in f_b, *i.e.*

$$C(c) \leqq C(b). \tag{2,16}$$

As $f(x)$ and $-f(x)$ have the same number of changes, we will suppose without any loss of generality that $a_n > 0$. If $a_n, \ldots, a_{n-k+1}$ and q are positive $a_{q,n-1} > a_{n-1}, \ldots, a_{q,n-k} > a_{n-k}$ and if q is great enough, $a_{q,n-k}$ becomes positive. In this manner we see that from a certain value d of q the coefficients become all positive. Hence $C(q) = 0$ for $q > c$. Therefore

Theorem. C decreases steadily to 0.

Let $b < c$ and let these values be not separated by a root of $f(x)$, then $a_{b,0} = f(b)$ and $a_{c,0} = f(c)$ have the same sign. Hence $C(b) = C(c) + 2k$, where $k \geqq 0$ is an integral number.

C is therefore a discontinuous decreasing function taking only integral values and the "saltus" at points which are not roots have even values. We have now to investigate the saltus in the roots of $f(x)$.

Let c be a root of multiplicity m, and $x - c = y$, then

$$f(x) = a_{c,n}y^n + \ldots + a_{c,m}y^m,\ a_{c,m} = 0.$$

Let $y = z + e$, $c + e = d$

$$f(x) = f_d(z) = a_{d,n}z^n + \ldots + a_{d,m}z^m + \ldots + a_{d,0}. \tag{2, 17}$$

Let $a_{c,m}y^m = a_{c,m}(z+e)^m = g(z)$.

If $e > 0$ there is no change in the coefficients of $g(z)$.

If $e < 0$, the coefficients have alternating signs and have therefore m changes. If e is small enough, the last $m+1$ coefficients of $f_d(z)$ have the same signs as $g(z)$. If therefore e increases from small negative to small positive values, the number of the changes in $a_{d,n}, \ldots, a_{d,m}$ decreases by an

even value, and the number of the changes in $a_{d,m}, \ldots, a_{g,s}$ decreases by m.

Hence C has at c a saltus $2l+m$, where $l \geqq 0$ is an integral number. By the help of these considerations we get the following theorem.

Theorem. Let $f(b) \neq 0$, $f(c) \neq 0$, $b<c$ and let r be the number of the roots of $f(x)$ in the interval $(0, c)$ every root being counted with its own multiplicity, then

$$C(b)=C(c)+r+2k, \qquad (2, 18)$$

where $k \geqq 0$ is an integral number.

Applying this theorem to an interval $(0\ c)$, where c is chosen so great that $C(c)=0$, we get as corollary

Descartes' rule. The number of the positive roots of $f(x)$ (every root being counted with its own multiplicity) is equal to the number of the changes of signs of the coefficients of $f(x)$ or to a number less than it by an even number.

If we consider that in (2, 12) $a_{s,s}$ and $f^{(s)}(b)$ have the same sign, (2, 18) can be expressed in the following manner.

Budan-Fourier's theorem. Let $f(b) \neq 0$, $f(c) \neq 0$, then the number of the roots of $f(x)$ in the interval (b, c) (every root being counted with its own multiplicity) is equal to the difference of the changes of signs in the sets

$$f(b), f'(b), \ldots, f^{(n)}(b) \text{ and}$$

$$f(c), f'(c), \ldots, f^{(n)}(c), \text{ or on an even number less than it}$$

If we set $x=\frac{cy+b}{y+1}$ and therefore $y=\frac{x-b}{c-x}$, then the positive roots of

$g(y)=(y+1)^n f(x)$ are in $(1, 1)$−correspondence to the roots of $f(x)$ in the interval (b, c). Therefore we are able to apply Descartes' rule to find out the number of the roots in this interval

These formulae do not always give directly the exact number of the roots in an interval, but they are very useful for getting it even in more complicated cases.

We will go into further details of the example considered in § 1.

$$f(x)=x^4-15x^3+68x^2-119x+67.$$

As we stated before, the real roots are positive and situated in the interval (1, 9). We could also get this result on considering the changes of signs. $f(-x)$ has no change and therefore no positive root, *i.e.* $f(x)$ has no negative root. From the previous calculations for this example we get — on considering the signs only—

$$C(0) = C(1) = 4$$
$$C(2) = 3$$
$$C(8) = 1.$$

We know two roots, one in the interval (1,2), another in the interval (8, 9); to each of these roots there corresponds a loss of one change. We should find out, if the loss of two changes in (2, 8) corresponds to roots of $f(x)$. For this purpose we try to approximate these suspected roots by Horner's scheme and get by very simple calculations

$$C(3) = 1 \qquad C(2{\cdot}6) = 3$$
$$C(2{\cdot}65) = 1 \qquad C(2{\cdot}64) = 3.$$

Hence the two roots can only be situated in the interval $2{\cdot}64 < x < 2{\cdot}65$, but we will prove that $f(x)$ is negative in this interval. We stated previously that

$$f(x) = 2 - 5(x-1) + 3(x-1)(x-2) + (x-1)(x-2)(x-3)(x-9)$$

$$= 2 - (x-1)\{5 - (x-2)[3 + (3-x)(9-x)]\}.$$ Hence for $2{\cdot}64 < x < 2{\cdot}65$

$$f(x) < 2 - 1{\cdot}64\{5 - 0{\cdot}65[3 + 0{\cdot}36 \,.\, 6{\cdot}36]\} < -0{\cdot}6612.$$

Hence $f(x)$ has only the two roots calculated in § 1. The same result can also be obtained by calculating the discriminant and stating that it is negative.

[2/3] A method to get the exact number of the different roots in any interval has been given by *Sturm*. We suppose that $(f(x), f'(x)) = 1$ and that therefore $f(x)$ has only simple roots. There is no loss of generality, viz. the h.c.f. of two polynomials can always be calculated by the algorithmus and if it should be of positive degree, $f(x)$ has to be replaced by $f(x) : (f(x)\, f'(x))$, which has the same roots but only simple ones.

The method uses a chain (*Sturm's chain*) of polynomials

$$f(x) = f_1(x), f_2(x), \ldots f_m(x), \qquad (2, 19)$$

and the number $C'(b)$ of the changes of sign in

$$f_1(b), \ldots, f_m(b).$$

The chain should be made in such a manner that $C'(b)$ is a steadily decreasing function changing its value only at the roots of $f(x)$, and having at these points a saltus of the value. Then

$$C'(b) - C'(c), \quad \text{for } b < c,\ f(b) \neq 0 \neq f(c) \tag{2, 20}$$

becomes the number of the roots of $f(x)$ in the interval (b, c).

For this purpose we have to arrange the chain so, that at each root of $f(x)$ one change will be lost and that in the roots of $f_{i>1}(x)$ the number of the changes will not be altered. In order to get a loss of changes in the roots of $f(x)$, f_1 must take the sign of f_2, when x passes a root of $f(x)$ from the left to the right, *i.e.*

(1) $f_2(x)$ has the same sign, as $f'(x)$.

In order to avoid an alteration in the value of C' at the roots of $f_2, \ldots, f_m$

(2) $f_m(x)$ should have constant sign, and

(3) for every root b of f_{i+1}, there is $f_i(b) \cdot f_{i+2}(b) < 0$.

So $f_{i+1}(x)$ will have the sign of exactly one of its neighbours before and after passing a root. The essence of these considerations is the following theorem.

Sturm's theorem. If the chain (2, 19) satisfies the conditions 1, 2, 3, the number of the roots of $f(x)$ in any interval (b, c) is given by (2, 20).

In order to get a chain of this kind we may use the algorithmus of the h. c. f.

$$f_2(x) = f'(x),\ f_i(x) = q_i(x) f_{i+1}(x) - f_{i+2}(x),$$

where the degrees of the polynomials decrease steadily to 0.

The first two conditions are obviously satisfied. As every common factor of $f_i(x)$ and $f_{i+1}(x)$ must be a common factor of $f_1(x) = f(x)$ and of $f_2(x) = f'(x)$, and as $(f(x), f'(x)) = 1$, $f_i(x)$ and $f_{i+1}(x)$ have no common root. If therefore $f_{i+1}(b) = 0$, $f_i(b) = -f_{i+2}(b) \neq 0$. Hence

$$f_i(b) \cdot f_{i+2}(b) < 0.$$

By Sturm's method it is always possible to get the exact number of the different roots in any interval, but the practical calculation is sometimes very burdensome, and it is often more convenient to use Budan-Fourier's theorem in connection with special considerations as we did it in the previous example. The reader may investigate the same example with Sturm's method as an *exercise*.

Remark. Sturm's chain (2, 19) can always be replaced by

$c_1 f_1(x), \ldots, c_m f_m(x)$ where $c_1, \ldots, c_m$ are positive constants.

[2/4] Sturm's theorem will now be applied on *Legendre's polynomials*

$$P_m(x) = \frac{1}{2^m \cdot m!} \cdot D^m[(x^2-1)^m] \; ; \quad m = 0, 1, 2, \ldots \tag{2, 21}$$

D^m denotes the m^{th} derivate of the function written in [ ], and D^0 is this function itself. If u and v are polynomials in x,

$$D^m(u.v) = \sum_{q=0}^{m} \binom{m}{q} D^{m-q}(u) D^q(v) \tag{2, 22}$$

$$D^m[(x^2-1)^m] = D^{m-1} D[(x^2-1)^m] = D^{m-1}[(x^2-1)^{m-1} . 2mx],$$

where from (2, 22) it follows for $m > 1$

$$D^m[(x^2-1)^m] = 2mx D^{m-1}[(x^2-1)^{m-1}] + 2m(m-1) D^{m-2}[(x^2-1)^{m-1}]. \tag{2, 23}$$

On the other hand we get from (2, 22) for $m > 1$

$$2D^m[(x^2-1)^m] = 2D^m[(x^2-1)^{m-1}(x^2-1)] = 2(x^2-1) D^m[(x^2-1)^{m-1}]$$

$$+ 4mx D^{m-1}[(x^2-1)^{m-1}] + 2m(m-1) D^{m-2}[(x^2-1)^{m-1}]. \tag{2, 24}$$

By subtraction of (2, 23) from (2, 24) and applying (2, 21) we get

$$m P_m(x) = (x^2-1) P'_{m-1}(x) + mx P_{m-1}(x). \tag{2, 25}$$

As $$P_1(x) = x, \quad P_0(x) = 1, \; P'_0(x) = 0,$$

(2, 25) holds also for $m = 1$, and therefore generally.

From

$$D^{m+1}[(x^2-1)^m] = D^m[2mx.(x^2-1)^{m-1}]$$

$$= 2mx D^m[(x^2-1)^{m-1}] + 2m^2 D^{m-1}[(x^2-1)^{m-1}]$$

we get

$$P'_m(x) = x P'_{m-1}(x) + m P_{m-1}(x). \tag{2, 26}$$

From (2, 25) and (2, 26) we eliminate P'_{m-1}, and get so

$$(x^2-1) P'_m(x) = mx P_m(x) - m P_{m-1}(x). \tag{2, 27}$$

On replacing m by $m+1$ in (2, 25) we get

$$(m+1)P_{m+1}(x) = (x^2-1)P'_m(x) + (m+1)xP_m(x). \qquad (2, 25')$$

From (2. 27) and (2, 25') we eliminate $P'_m(x)$ and we get

$$(m+1)P_{m+1}(x) = (2m+1)xP_m(x) - mP_{m-1}(x). \qquad (2, 28)$$

We consider the sequence

$$P_n(x).\ P_{n-1}(x),\ldots,P_1(x).P_0(x)=1 \qquad (2, 29)$$

in the interval

$$-1 \leqq x \leqq +1.$$

From (2. 27) it follows for $m=n$, that if $P_n(x)=0$, $P'_n(x)$ has the same sign as $P_{n-1}(x)$. If $P_{m+1}(x)$ and $P_m(x)$ would have a common root, this root has to be a root of $P_{m-1}(x)$ too, as we see from (2, 28) and therefore of all subsequent polynomials of (2, 29), in contradiction to $P_0(x)=1$. Hence for $P_m(x)=0$, $P_{m+1}(x) \neq 0$, and therefore it follows from (2, 29) that $P_{m+1}(x)P_{m-1}(x)<0$ at every root of $P_m(x)$.

As $P_n(x)$ and $P_{n-1}(x)$ have no common root, it follows from (2, 27) that $P_n(x)$ has no common root with its derivate and has therefore simple roots only. Hence (2, 29) is a chain of Sturm in the interval $-1 \leqq x \leqq +1$ for every n.

From (2, 27) it follows that

$$P_m(1) = P_{m-1}(1)$$

and

$$P_m(-1) = -P_{m-1}(-1).$$

As $P_0(x)=1$, it follows that

$$P_m(1) = 1$$

and

$$P_m(-1) = (-1)^m.$$

The number of changes of sign (2,29) is therefore $C'(-1)=n$, $C'(1)=0$. Hence $P_n(x)$ has n different roots in the interval $(-1,\ 1)$. From (2, 21) it follows that $P_n(x)$ is of degree n. Hence the roots of Legendre's polynomials are all situated in the interval $(-1, +1)$ and are simple roots.

To find out systematically the real roots of a polynomial [2/5]

$$f(x) = a_0 + a_1x + \ldots + a_nx^n$$

with real coefficients, we have at first to find an interval containing all these roots.

Let

$$t=1+\left|\frac{a_{n-1}}{a_n}\right|+\ldots+\left|\frac{a_1}{a_n}\right|+\left|\frac{a_0}{a_n}\right|.$$

As we stated in Part II, § 13, $f(x)$ has the same sign as a_n, and therefore $f(x)$ has the same sign as $(-1)^n a_n$, if $x \geqq t$. Hence the roots of $f(x)$ are situated in the interval $(-t, +t)$.

When an interval containing all the real roots of the polynomial has been found, we subdivide it and find by Sturm's method how many roots are contained in the subintervals. The subintervals containing roots should be subdivided again, and the procedure must be repeated till an interval (a, c) has been found for every root b so that $a<b<c$, and $c-a$ is less than the approximation required for the special problem. As the root should mostly be represented by a decimal fraction, the first subdivision will mostly be made by the integral values of x, the second subdivision by numbers with one decimal, etc., then after $m+2$ subdivisions the root will be determined up to the m^{th} decimal.

[2/6] This method to calculate the roots will seldom be used by a clever reckoner. By the practice of reckoning he will get some idea what subinterval may contain roots, and he will therefore calculate the polynomial only for the endpoints of those intervals. In an interval containing only one simple root, the approximation mostly comes out very quickly by Horner's scheme.

Regula falsi. If $f(a)$ and $f(b)$ have different signs, the graph of the function $f(x)$ may be replaced by the straight line connecting the points with abscisses a and b. This line intersects the x-axis in $x=(a-b).f(a):[f(b)-f(a)]$. This value may be considered as a first approximation of the root. Let us consider the example of § 1.

$$x-1=y,\ f(x)=g(y)=y^4-11y^3+20y^2-24y+2$$

$$g(0)=2,\ g(1)=-3.$$

Hence we get by the *regula falsi* an approximation $y_1=0{\cdot}4$. We will try to improve this approximation by

$$24y^2=y_1{}^4-11y_1{}^3+20y_1{}^2+2 \backsim 6. \qquad (2, 30)$$

Hence

$$y_2 \backsim 1:4.$$

This approximation is better, but it is not good. As we calculated in § 1.

$$x-1=0{\cdot}0935324\ldots$$

Of course the graph of the polynomial is in that interval very different from a straight line. Now

$$g(0)=2 \qquad g(1)=-3$$

$$g'(0)=-24 \qquad g'(1)=5.$$

The graph is therefore considerably bent in the interval, and the root must obviously lie near the point $x-1=0$.

Newton's method. The graph becomes approximated by the tangent; *i.e.* the terms of higher degree in Horner's scheme have to be omitted. This method gives good results, when the distance from the root is small, the term of 1st degree has a coefficient of high absolute value and the absolute values of the coefficients of the higher terms are not too big. This method has been used in the above example, and gave the approximation $2:24 \backsim 0{\cdot}1$. We can improve this solution by using the equation (30) with $y_1=0{\cdot}1$ then y_2 becomes $<0{\cdot}095$.

As for $y_1=0, \qquad y_2>0{\cdot}082$

and for $y_1=0{\cdot}1, \qquad y_2<0{\cdot}095$ hold,

and as y_2 is a continuous function of y_1, there must be in the interval $(0{\cdot}082, 0{\cdot}095)$ a value for which $y_1=y_2$, and that is a root. So Newton's method is very useful in certain cases, but if the interval is big or the tangent makes only a small angle with the axis, the method cannot be used.

The suitable choice of the methods should be learned by practice. In this section sometimes reference has been made to the graph of a polynomial. So the reader may ask if *graphical methods* may not be helpful to get the roots of a polynomial. Of course methods of this kind exist and are very helpful to get a convenient first approximation for the roots, but in applying these methods, only those readers may succeed, who are familiar with the theory and practice of mathematical drawing.*

Let a be a fixed real number $\neq 0$, and let b and c be two different and consecutive roots of $f(x)$. then $f(x)$ has a constant sign in the interval [2/7]

* A very useful graphical method is, *e.g.*, Lill's rectangular method. For reference see Bieberbach-Bauer, Vorlesungen über Algebra, pp. 134-140.

(b, c). Let this sign be $+$, then $f(x)$ is increasing in an interval (b, b'), and therefore $f'(x)$ is non-negative in this interval ; we can choose b' in such a manner that in the points $x \neq b$. $g(x) = a f(x) + f'(x) > 0$, *viz.*, either $f'(b) > 0$, or $f'(x)$ becomes zero of a smaller order than $f(x)$. In the same manner we can find out an interval (c', c), where $g(x) < 0$ for $x \neq c$. Hence $g(x)$ has an odd number of roots in the interior of the interval (b, c). The same holds if $f(x)$ is negative in (b, c).

If b is a root of $f(x)$ of multiplicity $r+1$, it is a root of $g(x)$ of multiplicity r. Let m_1 be the number of the different real roots of $f(x)$ and m_2 the number of these roots counted with their multiplicity, then the corresponding numbers m'_1, and m'_2 for $g(x)$ satisfy $m'_1 \geqq m_1 - 1$

$$m'_2 \geqq m_2 - 1.$$

But $n - m_2$ is the number of the complex roots of $f(x)$, and therefore even, and as $g(x)$ is of degree n, the number $n - m'_2$ is even, and therefore $m'_2 \geqq m_2$. If $m'_1 = m_1 - 1$, there would be in every interval (b, c) between different roots exactly one root of $g(x)$, and one of them would be a double root. That is impossible, *viz.*, the number of the roots of $g(x)$ in such an interval should be odd, when every root is counted with its own multiplicity. Hence $m'_2 \geqq m_2$. So we get the

Lemma. Let $a \neq 0$, then the number of different real roots of $a f(x) + f'(x)$ is not less than the number of different real roots of $f(x)$, and the corresponding proposition holds, when each root has been counted with its multiplicity.

This Lemma is a special case of the following theorem.

Poulain's theorem. Let $h(z) = b_m z^m + \ldots + b_1 z + b_0$ be a polynomial with real coefficients and real roots only, let $b_m \neq 0$, $b_0 \neq 0$ and let $f(x)$ be a polynomial with real coefficients, then

$$g(x) = b_0 f(x) + b_1 f'(x) + \ldots + b_m f^{(m)}(x)$$

has not less different real roots than $f(x)$, and the corresponding proposition holds, when each root has been counted with its own multiplicity.

Proof. Without any loss of generality we suppose that $b_m = 1$. Hence for $m = 1$, $g(x) = b_0 f(x) + f'(x)$, and in this case the theorem becomes reduced to the lemma. Let $p > 0$, and let the theorem be proved for $m = p$. We will prove it for $m = p + 1$. If a is a root of $h(z)$, then $a \neq 0$, and $h(z) = (z - a) \cdot h_1(z)$, where $h_1(z)$ has real roots only.

Let $h_1(z) = z^p + \ldots + c_1 z + c_0$. As the theorem holds for $m = p$, the number of the roots of $g_1(x) = \sum_{i=0}^{p} c_i f^{(i)}(x)$ is greater than or equal to the number of the roots of $f(x)$. In this formula $f^{(0)}(x)$ means $f(x)$.

$$g'_1(x) = \sum_{i=0}^{p} c_i f^{(i+1)}(x).$$

If we replace in $h(z) = (z-a)\, h_1(z)$ the powers of z by the corresponding derivates of $f(x)$, we get therefore

$$g(x) = g'_1(x) - a\, g_1(x).$$

From the preceding lemma it follows that the number of the roots of $g(x)$ is not less than the number of the roots of $g_1(x)$, and therefore not less than the number of the roots of $f(x)$. Hence the theorem holds.

§ 3. Graeffe's Method.

By *Graeffe's method* all the roots can be calculated without preparatory measures at the same time. The method leads to suitable results very quickly, when the roots are all different and real. It is mostly difficult to estimate exactly the error made by suitable omissions. Hence the results have to be examined by putting in the equation. [3/1]

Let
$$b_1 > b_2 > \ldots > b_n \qquad (3, 1)$$
be the roots of the polynomial $a_0 x^n + a_1 x^{n-1} + \ldots + a_n$, then

$$\frac{-a_1}{a_0} = b_1\left(1 + \frac{b_2}{b_1} + \ldots + \frac{b_n}{b_1}\right) = b_1 (1 + \epsilon_1)$$

$$\frac{-a_2}{a_1} = \frac{-a_2}{a_0} : \frac{a_1}{a_0}$$

$$= \frac{b_1 b_2\left(1 + \frac{b_3}{b_1} + \ldots + \frac{b_n}{b_1} + \frac{b_3}{b_2} + \ldots + \frac{b_n}{b_2} + \frac{b_3 b_4}{b_1 b_2} + \ldots + \frac{b_{n-1} b_n}{b_1 b_2}\right)}{b_1(1+\epsilon_1)}$$

$$= b_2 (1 + \epsilon_2)$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$-\frac{a_{n-1}}{a_{n-2}} = b_{n-1}(1 + \epsilon_{n-1})$$

$$\frac{-a_n}{a_{n-1}} = (-1)^n \cdot \frac{a_n}{a_0} : \left(\frac{-a_1}{a_0} \cdot \frac{-a_2}{a_1} \cdot \ldots \cdot \frac{-a_{n-1}}{a_{n-2}}\right)$$

$$= \frac{b_1 \ldots b_n}{b_1(1+\epsilon_1) \ldots b_{n-1}(1+\epsilon_{n-1})} = b_n(1 + \epsilon_n)$$

If $b_i : b_{i-1}$ is very great, for $i = 1, \ldots, n$ the numbers e_i can be omitted. In this case we get the approximation

$$b_i \backsim \frac{-a_i}{a_{i-1}} \qquad \text{for } i = 1, \ldots, n. \tag{3, 2}$$

In general (3, 2) is not a consequence of (3, 1), but for a suitable exponent m the quotients $b_{i-1}^m : b_i^m$ become negligible. Therefore we have to find out a polynomial, whose roots are $b_1^m, \ldots, b_n^m$. The coefficients of this polynomial are symmetric functions of $b_1, \ldots, b_n$, hence it is possible to calculate them as rational functions of $a_0, a_1, \ldots, a_n$ with rational coefficients. The calculation for an arbitrary m is tiresome, but it is easy to find out a polynomial whose roots are the squares of the roots of $f(x)$ and by repeating this construction we get subsequently polynomials with the roots

$$b_i^2, b_i^4, b_i^8, \ldots, b_i^{2^k} \ldots$$

Let $a'_0 x^n + a'_1 x^{n-1} + \ldots + a'_n$, $a'_0 = a_0^m$ have the roots $b_1^m, \ldots, b_n^m$, and let m be chosen so great, that $b_{i+1}^m : b_i^m$ can be omitted, then $b_i^m \backsim \frac{-a'_i}{a'_{i+1}}$ holds.

The corresponding holds for the polynomial with the roots $b_1^{2m}, \ldots, b_n^{2m}$; hence the absolute values of its coefficients become approximately the squares of the corresponding coefficients a'_i. Hence we have to repeat the construction of polynomials till the calculation shows that after further repetition, the coefficients will become practically the squares of the coefficients of the preceding polynomial. To get a polynomial whose roots are the squares of the roots of $f(x)$, we calculate

$$\begin{aligned}(-1)^n f(x) \cdot f(-x) &= a_0(x-b_1) \cdot \ldots \cdot (x-b_n)\, a_0(x+b_1) \cdot \ldots \cdot (x+b_n) \\ &= a_0^2(x^2-b_1^2) \cdot \ldots \cdot (x^2-b_n^2) \\ &= f_2(x^2).\end{aligned}$$

The coefficients of f_2 will be calculated by the following scheme

(1)	a_0	a_1	a_2	$\ldots$	a_n
	a_0	$-a_1$	a_2	$\ldots$	$\pm a_n$
(2)	a_0^2	$-a_1^2$	a_2^2	$\ldots$	$\pm a_n^2$
		$+2a_1a_2$	$-2a_1a_3$		
			$+2a_0a_4$		

As in the first pair of lines corresponding numbers differ only by the sign. it is usual to write only the signs in the 2^d line. The numbers increase very quickly ; therefore it is convenient to omit the last figures denoting the decimals very clearly. For this purpose we shall use the notation

$3^9456131$ for $3{\cdot}456131 \cdot 10^9$.

To extract the roots at the end of the calculation we need logarithms. It is therefore useless to calculate more decimals than the tables of logarithms contain.

Example. $x^3 - 10x^2 + 16x - 2 = 0$

(1)	1	$-1^{1}0$	$1^{1}6$	-2
	$+$	$+$	$+$	$+$
	1	$-1^{2}0$	$2^{2}56$	-4
		$+0\ 32$	$-0\ 4$	
(2)	1	$-6^{1}8$	$2^{2}16$	-4
	$+$	$+$	$+$	$+$
	1	$-4^{3}624$	$4^{4}6656$	-16
		$+0\ 432$	$-0\ 0544$	
(4)	1	$-4^{3}192$	$4^{4}6112$	-16
	$+$	$+$	$+$	$+$
	1	$-1^{7}75729$	$2^{9}12631$	-256
		$+0\ 00932$	$+0\ 00013$	
(8)	1	$-1^{7}74797$	$2^{9}12618$	-256

In the next step the coefficients will become the squares of the preceding coefficients, and in no case the error will have influence on the first 5 figures. Therefore we stop the procedure, and calculate now the roots by the help of logarithms.

log. of the coefficients	log. x^8	log. $\lvert x \rvert$	$\lvert x \rvert$
0	7·24254	0·90532	8·0412
7·24254	2·08506	0·26063	1·8223
9·32760	1·08064—8	0·13508 − 1	0·1365
2·40224			
		0·30103	10·0000
		= log 2	

The sign of the roots cannot be determined by Graeffe's method ; we have to arrange a special investigation for the signs in every case. In this example the coefficients have alternating signs, hence the coefficients of $f(-x)$ are all positive. The real roots of $f(-x)$ are therefore negative hence the roots of $f(x)$ are all positive.

For *checking* we form the elementary symmetric functions of the approximate roots, and we get

$$s_1 = 10, \qquad s_2 = 15{\cdot}9998 \qquad s_3 = 2$$

for 10 16 2 .

[3/2] If a real polynomial has complex roots, two of them are always conjugate, and these have therefore the same absolute value. Graeffe's method has therefore to be modified in this case. An example will give valuable hints for necessary modifications.

Example. $x^4 - 11x^3 + 29x - 24x + 2$.

We know from § 1 that this polynomial has two real roots $b_1 = 7{\cdot}5925 \ldots$ and $b_4 = 0{\cdot}095324 \ldots$ and two complex roots.

The calculation by Graeffe's method is given * on the next page.

If the procedure be repeated, the two first and the two last coefficients will become the squares of the corresponding coefficients of the line (8), but the third coefficient will depend also upon the second and the fourth. We cannot expect that further repetition of the procedure will make the third coefficient independent of his neighbours, as two roots of the polynomial have an equal absolute value. If b_1 is greater than the absolute value of

* As the sign in the 2^4 line is always + we omit these lines for abbreviation.

	1	− 11	29	− 24	2
(1)	1	− 121	841	− 576	4
		58	− 528	+ 116	
			+ 4		
(2)	1	− 63	317	− 460	4
	1	− 3969	100489	− 211600	16
		+ 634	− 57960	+ 2536	
			8		
(4)	1	− 3335	42537	209064	16
	1	$-1^{7}11222$	$1^{9}80938$	$-4^{10}3707$	256
		+ 851	− 1 41525	0	
			0		
(8)	1	$-1^{7}10371$	$3^{8}9413$	$-4^{10}3707$	256 .

the complex roots, then

$$\frac{-a'_1}{a'_0} = b_1^{m}\,\frac{(1 + 2b_2^{m} + b_4^{m})}{b_1^{m}} \backsim b_1^{m} \text{ for a suitable } m .$$

A rough mental calculation shows that $b_1^{2} \backsim 60$, $b_1^{8} \backsim 1^{7}2$.

The same consideration made for $f(\frac{1}{x})$ shows that, if $b_4 < b_2$.

$\frac{-a'_4}{a'_3} \backsim b_4^{m}$ holds. Hence the complex roots are only dependent on the 3 middle coefficients. In order to get the law of dependence we shall generalise the considerations.

Let B and C be two intervals so that every number of C is very small in comparison to the numbers of B, and let

$$f(x) = a_r + a_{r+1}\,x + \ldots + a_s\,x^{n}, \quad n = r + s$$

have two sets of roots

$b_1, b_2, \dots, b_r$ whose absolute values belong to B, and

$c_1, c_2, \dots, c_s$ " " " " " C.

Let $m \leqq r$, then $\frac{(-1)^m a_m}{a_0}$ becomes approximately equal to the m^{th} symmetric fundamental function of $b_1, \dots, b_r$, and

$$\frac{a_{r+1}}{a_0} = \frac{a_r}{a_0} \cdot \left(\frac{s}{t}\right) z_t^{\,r},$$

where z_t is a suitably chosen mean-value of the roots of the second set, and therefore a number of C. Let y be a number of B, then

$$\frac{f(y)}{a_r y^r} = \frac{a_0}{a_r} y^r + \dots + 1 + \frac{a_{r+1}}{a_r} y^{-1} + \dots + \frac{a_s}{a_r} y^{r-s}$$

$$\backsim \frac{1}{a_r} \sum_{i=0}^{i=s} a_i y^{r-i} = \frac{1}{a_r} f_1(y).$$

Let $a_r = 1$, and y be one of the roots b_t, then $f_1(y) \backsim 0$, and we can therefore approximate the 1st set of the roots of $f(x)$ by the roots of $f_1(x)$. But, as a common factor of the coefficients has no importance, the roots b_t are approximated by the roots of

$$a_0 x^r + \dots + a_r .$$

Let $x = \frac{1}{z}$, then $a_0 z^r + \dots + a_1 z + a_r$

has the roots $\frac{1}{c_1} = b'_1, \dots, \frac{1}{c_s} = b'_s$,

and $\frac{1}{b_1} = c'_1, \dots, \frac{1}{b_r} = c'_r$;

the absolute values of b'_t belong to an interval B', the absolute values of c'_t belong to C' and every number of C' is small in comparison to B'. Hence the roots b'_t can be approximated by the roots of $a_0 z^r + \dots + a_r$, and therefore the roots c_t of $f(x)$ can be approximated by the roots of

$$a_r x^r + a_{r-1} x^{r-1} + \dots + a_0 .$$

So the polynomial $f(x)$ has to be split up in two polynomials, the first is defined by the $r+1$ upper terms and leads to the upper class of roots,

the second one is defined by the $s + 1$ lower terms, and leads to the lower class of roots.

The two classes may also be divided into sub-classes etc. Finally we get classes

$$b_{1\cdot 1}, \dots, b_{1\cdot r_1}\ ;\ b_{2\cdot 1}, \dots, b_{2\cdot r_2}\ ;\ \dots\ ;\ b_{s\cdot 1}, \dots, b_{s\cdot r_s},$$

each root being small in comparison with the roots of the preceding classes, and to each class corresponds a polynomial, which can be cut out from $f(x)$. The ratio of the absolute value of the roots increases when we replace these roots by higher powers of them, therefore we get finally by Graeffe's method k polynomials each of them having only roots with the same absolute value. In the previous example these polynomials are

$$x - 1^{7}10371, \quad 1^{7}10371\cdot x^2 - 2^{8}9413\, x + 4^{10}3707, \quad 4^{10}3707\, x - 256.$$

From these polynomials we get the roots of $f(x)$

$8 \log |\, b_1 \,| = 7{\cdot}04286$ $\qquad \log |\, b_1 \,| = 0{\cdot}88036$ $\qquad |\, b_6 \,| = 7{\cdot}592$

$8 \log |\, b_4 \,| = 7{\cdot}76769 - 16$ $\qquad \log |\, b_4 \,| = 0{\cdot}97096 - 2$ $\qquad |\, b_4 \,| = 0{\cdot}093532$

$16 \log |\, b_2 \,| = \log 4^{10}3707 - \log 1^{7}10371$

$= 3{\cdot}59767$ $\qquad \log |\, b_2 \,| = 0{\cdot}22476$ $\qquad |\, b_2 \,| = 1{\cdot}6779$

$\log \cos 8\phi = \log 3^{8}9413 - 8 \log |\, b_2 \,| \quad - \log 2 - \log 110371 \quad = 9{\cdot}45293 - 10$

$8\phi = +73°31' \ + k\cdot 360°$

$\phi = \pm 9°11'23'' + k\cdot 45°$

To finish this calculation we have to fix the signs of the real roots and to determine the integral number k. As the signs of the coefficients are alternating, there is no negative root. Hence $b_1 = 7{\cdot}592, \dots, b_4 = 0{\cdot}093532\dots$. These numbers correspond to the results obtained in § 1 by Horner's method and by Lagrange's method.

As $b_1 + b_2 + b_3 + b_4 = 11$, $2r\cos\phi = 3{\cdot}315\dots$. But as $2r = 3{\cdot}356\dots$, ϕ must be a very small angle. Hence $k = 0$. So we get

$$b_2 = 1{\cdot}6567 + i\,0{\cdot}26803$$

$$b_3 = 1{\cdot}6567 - i\,0{\cdot}26803.$$

For checking : $\log b_1 + \log b_4 + 2 \log r = 0{\cdot}30104$

for $\log 2 = 0{\cdot}30103$

$b_1 + b_2 + b_3 + b_4 = 10{\cdot}999$

for 11.

If we replace in this last verification the value for b_1 by the more exact value obtained from § 1 $b_1 = 7{\cdot}52925$ we will get

$b_1 + b_2 + b_3 + b_4 = 10{\cdot}9994$. The result can be corrected by further calculation. As we see from the results of § 1 and from the checking given here b_1, b_4, and r are very exact. The correction is therefore expected to concern mainly the angle ϕ, whose true value may be a little smaller. As ϕ itself is a small angle, this correction will materially affect $\sin \phi$. Hence the imaginary parts of b_2 and b_3 are true up to the second decimal only.

If a polynomial with roots of equal absolute value has a degree > 2, either it has multiple roots, or it has non-conjugate roots. The multiple roots will be removed, when we divide by the h. c. f. of the polynomial and its derivative. Non-conjugate roots of equal absolute value can be cleared away by Horner's scheme, *viz.*, if $| x | = | x' |$ and x' is different from x and $\bar{x}$, then $| x - a | \neq | x' - a |$.

Hence the real and the complex roots of $f(x)$ can be found out by a combination of Graeffe's method and Horner's scheme in every case. The results should be verified and it is possible to minimise the error by the methods given in § 1.

§ 4. Roots of complex polynomials.

Let $\phi(x)$ be a polynomial with complex coefficients,

$$\phi(x) = a_0 + a_1 x + \ldots + a_n x^n$$

$$\bar{\phi}(x) = \bar{a}_0 + \bar{a}_1 x + \ldots + \bar{a}_n x^n,$$

$(\phi(x), \bar{\phi}(x)) = f_1(x)$, $\phi(x) = f_1(x)\, \bar{\phi}_1(x)$, $\phi(x) = f_1(x)\, \bar{\phi}_1(x)$

$\phi_1(x)\, \bar{\phi}_1(x) = f_2(x)$,

then $f_1(x)$ and $f_2(x)$ are real polynomials. On applying Graeffe's method to these polynomials, we get the roots of $\phi(x)$, but out of two conjugate

roots of $f_2(x)$ one is a root of $\phi(x)$, the other is a root of $\bar{\phi}(x)$, and we have therefore to make a verification finally.

Let $|z| \geqq \sum_{j=0}^{n} \left|\frac{a_j}{a_n}\right| = t$, then

$$\left|\frac{1}{a_n}\right| |\phi(z)| \geqq \Big| |z^n| - \sum_{k=0}^{n-1} \left|\frac{a_k}{a_n}\right| |z|^k \Big| \geqq \Big| |z|^n - |z|^{n-1} \quad (|z|-1) \Big| = |z|^{n-1} \quad > 0$$

hence $\phi(t) \neq 0$, and therefore the absolute value of the roots of $\phi(x)$ is $< t$.

Another limit for the roots can be found out by Kakeya's theorem.

The roots of $\phi(x)$ are also roots of the real polynomial $f_1(x)$. $f_2(x) = f(x')$, where $x' = x + a$ and the real number a can be chosen in such a manner that $f(x') = a_n x'^n + \ldots + a_1 x' + a_0$ has positive coefficients only.

For the polynomials with positive coefficients the following theorem holds.

Theorem. Let the coefficients of $f(x) = a_0 + a_1 x + \ldots + a_n x^n$ be positive and $0 < p < \frac{a_{k-1}}{a_k} < q$ for $k = 1, \ldots, n$, then the roots of $f(x)$ have to satisfy the condition

$$p < |x| < q.$$

Proof. Let $x = q\,y$, $\quad f(x) = g(y) = \sum b_k y^k$, then $b_k = q^k a_k$.

Hence $b_{k-1} : b_k < 1$. From Kakeya's theorem it follows therefore, for the roots that $|y| < 1$, and $|x| < q$.

The roots of $F(z) = a_0 z^n + \ldots + a_{n-1} z + a_n$ are reciprocal to the roots of $f(x)$. As $\frac{a_{n-k+1}}{a_{n-k}} < \frac{1}{p}$ holds, it follows from the first part of the proof that the roots of F have to satisfy

$|z| < \frac{1}{p}$. Hence $|x| = \left|\frac{1}{z}\right| > p$ holds for the roots of $f(x)$.

An interesting connection between the roots of $\phi(x)$ and its derivate $\phi'(x)$ is given by the

Theorem of Gauss. Every convex polygon including all the roots of $\phi(x)$ contains every root of $\phi'(x)$.

Proof. Without any loss of generality we can suppose that ϕ and ϕ' have no common root. Let γ be an arbitrary root of ϕ' and $\beta_1, \dots, \beta_n$ be the roots of ϕ, then

$$\frac{\phi'(x)}{\phi(x)} = \Sigma \frac{1}{x-\beta_i}, \quad \text{hence } 0 = \frac{\phi'(\gamma)}{\phi(\gamma)} = \Sigma \frac{1}{\gamma-\beta_i}, \text{ and therefore}$$

$$0 = \Sigma \frac{1}{\bar{\gamma}-\bar{\beta}_i} = \Sigma \frac{\gamma-\beta_i}{|\gamma-\beta_i|^2} = \Sigma (\gamma-\beta_i) \cdot b_i \quad \text{where } b_i \text{ is positive.}$$

We consider the geometrical representation of the complex numbers in the plane. $(\gamma-\beta_i)\, b_i$ are vectors starting from γ and directed to $\beta_1, \dots, \beta_n$. As the sum is equal to 0, every component of this sum is equal to 0. Let G be an arbitrary straight line passing through γ. The components of $(\gamma-\beta_i)\, b_i$ orthogonal to G form a sum equal to zero, hence either the components are all equal to zero, or there are components with different sign. In the 1st case the points β_i are all situated on G; in the 2nd case, there are roots of ϕ on both sides of G. In no case there are roots of ϕ on one side of G only. Let now P be a convex polygon including all the roots of ϕ. If γ is outside of P, we can draw a straight line G not intersecting P through γ. Hence P and therefore all the roots of ϕ' are situated on the same side of G. Hence γ is not a root of ϕ'.

Let P_0 be the smallest convex polygon including the roots of ϕ. (The reader may prove that such a polygon exists and is unique.) P_1 the corresponding polygon defined by $\phi', \dots$, P_i the smallest polygon containing the roots of $\phi^{(i)}$. The polygons with higher indices are included in the preceding. $\phi^{(n)}$ degenerates to the point $\frac{-1}{n} \cdot \frac{a_{n-1}}{a_n} = \frac{1}{n} \Sigma a_i$. This point is the centre of gravity of the roots of ϕ, and for the same reason it is the centre of gravity of the roots of ϕ', and of the roots of each derivate.

§ 5. Interpolation.

[5/1] Let

$$\beta_1, \dots, \beta_{n+1} \tag{5, 1}$$

be $n+1$ *different* elements of an arbitrary field K, and let

$$\lambda_1, \dots, \lambda_{n+1} \tag{5, 2}$$

be $n+1$ arbitrary elements of K.

We want to find out a polynomial $f(x)$ of K $[x]$ so that

$$f(\beta_i)=\lambda_i \text{ for } i=1, \dots, n+1, \quad \text{and degree } f(x)\leqq n.$$

Let $f(x)=a_0+\dots+a_nx^n$. This polynomial has the proposed properties if and only if its coefficients satisfy

$$\sum_{i=0}^{n} a_i \beta_k^i = \lambda_k.$$

The determinant of this system of $n+1$ linear equations (see Part II [10/4]) is equal to $\pm\prod_{i<j}(\beta_i-\beta_j)$ and is $\neq 0$, as the $n+1$ elements β_i are supposed to be different. Hence the problem has one and only one solution. This solution can be calculated by the methods explained in Part I, but it is easier to get it from special cases.

Let $f_k(x)$ be the solution if $\lambda_{i\neq k}=0$, $\lambda_k=1$, then $f(x)=\sum_{k=1}^{n+1} \lambda_k f_k(x)$ is the solution for arbitrary λ-elements. But $f_k(x)=\dfrac{g(x)}{(x-\beta_k)\ g'(\beta_k)}$, where $g(x)=\prod_{i=1}^{n+1}(x-\beta_i)$. So we get *Lagrange's formula for interpolation.*

$$f(x)=\Sigma g(x)\frac{\lambda_k}{(x-\beta_k)\ g'(\beta_k)}. \tag{5, 3}$$

By Lagrange's formula the problem of interpolation has been solved in the most complete and general manner, but the formula is not convenient for practical calculation. It is easier to calculate the coefficients of the product representation of $f(x)$ [5/]

$$f(x)=\gamma_0+\gamma_1(x-\beta_1)+\gamma_2(x-\beta_1)(x-\beta_2)+\dots+\gamma_n(x-\beta_1)\dots(x-\beta_n). \tag{5, 4}$$

Here is $\gamma_0=f(\beta_1)=\lambda_1$, $\gamma_1=\dfrac{\lambda_2-\lambda_1}{\beta_2-\beta_1}$, and we may successively calculate the coefficients γ_i. It is convenient to arrange this calculation in the following manner.

Let $f_i(x)$ be defined by $f_0(x)=f(x)$, and for $k=1, \dots, n$

$$f_k(x)=\frac{f_{k-1}(x)-f_{k-1}(\beta_k)}{x-\beta_k};$$

then $f_k(x)=\gamma_k+\gamma_{k+1}(x-\beta_{k+1})+\dots+\gamma_n(x-\beta_{k+1})\dots(x-\beta_n)$.

Hence $f_i(\beta_{i+1})=\gamma_i$. We have therefore to calculate the values

$$\{k, m\}=f_k(\beta_m) \text{ for } k=0\ldots n, \qquad k<m\leqq n+1$$

by $$\{k, m\}=[\{k-1, m\}-\{k-1, k\}] : (\beta_m-\beta_k)$$

and $\{0, m\}=\lambda_m$. We calculate the values column-wise in the following scheme

	$\{0, m\}$	$\{1, m\}$	$\{2, m\}$	...	$\{n, m\}$
$\{k, 1\}$	λ_1				
$\{k, 2\}$	λ_2	$\frac{\lambda_2-\lambda_1}{\beta_2-\beta_1}$			
$\{k, 3\}$	λ_3	$\frac{\lambda_3-\lambda_1}{\beta_3-\beta_1}$	$\frac{\{1, 3\}-\{1, 2\}}{\beta_3-\beta_2}$		
.	.				
.	.				
.	.				
$\{k, n+1\}$	λ_{n+1}	$\frac{\lambda_{n+1}-\lambda_1}{\beta_{n+1}-\beta_1}$	$\frac{\{1, n+1\}-\{1, 2\}}{\beta_{n+1}-\beta_2}$	...	$\frac{\{2-1, n+1\}-\{n-1, n\}}{\beta_{n+1}-\beta_n}$

The first elements of the different columns of this scheme form the set $\gamma_0, \gamma_1, \ldots, \gamma_n$ of the coefficients of (5, 4). This scheme is easier for calculation than Lagrange's formula.

[5 /3] The reckoning can further be simplified if the elements $\beta_1, \ldots, \beta_{n+1}$ are equidistant, *i.e.*, if

$$\beta_{k+1}-\beta_k=\Delta_x$$

for every k ; then

$$\Delta_x \,.\, \{k, m\}=[\{k-1, m\}-\{k-1, k\}] \; : (m-k)$$

$$=\frac{1}{m-k} \sum_{i=k}^{m-1} \Delta_{k-1,\, i} \,,$$

where $\Delta_{k-1,\, i}=\{k-1, i+1\}-\{k-1, i\}$ is the difference of two consecutive elements in the preceding column. So $\Delta\ \{k, m\}$ is the *mean-value* of the differences of consecutive elements of the rows m to k in the column $k-1$.

We will now transform the scheme for these cases in such a manner that we have to calculate the differences of consecutive elements only and not the mean-values.

For this purpose we introduce the notations usual in the calculus of differences.

Let $\Delta_x \,.\, u = x - \beta_1$, then $\Delta_x(u-k-1) = x - \beta_{k+1}$.

Let $F(u) = f(x) = f(\Delta_x u + \beta_1) = \gamma_0 + \gamma_1 u \Delta_x + \gamma_2 u \,.\, (u-1)\Delta_x^2 + \ldots + \gamma_n u \,.\, (u-1)\ldots(u-n-1)\Delta_x^n$

and let $\Delta f(x) = f(x+\Delta_x) - f(x) = F(u+1) - F(u)$, then

$\Delta f(x) = \Delta_x \,.\, [\gamma_1 + 2\gamma_2 u \Delta_x + \ldots + n\gamma_n u.(u-1)\ldots(u-n-2).\ \Delta_x^{n-1}]$

viz., $(u+1)u(u-1)\ldots(u-k+1) - u(u-1)\ldots(u-k) = (k+1)u(u-1)\ldots(u-k+1)$.

Let $\Delta(\Delta f(x)) = \Delta^2 f(x)$, ... , $\Delta(\Delta^r f(x)) = \Delta^{r+1} f(x)$;

then we get by repetition of this procedure

$\Delta^2 f(x) = \Delta_x^2 [2\gamma_2 + 2.3\gamma_3 . u\Delta_x + \ldots + n(n-1)\gamma_n u(u-1)\ldots(u-n-3)\Delta_x^{n-2}]$

$\Delta^n f(x) = \Delta_x^n \,.\, n\,!\,\gamma n.$

For abbreviation we shall write $\Delta^i f(\beta_i) = \Delta_i^i$. Then

$$\Delta_i^{i+1} = \Delta_{i+1}^i - \Delta_i^i$$

and

$$f(\beta_1) = \gamma_0$$

$$\Delta_1^1 = \Delta_x \gamma_1$$

.....

$$\Delta_1^i = \Delta_x i\,!\,\gamma_i$$

.....

$\Delta_1^n = \Delta_x^n n\,!\,\gamma_n$ (for $i = 1, 2\ldots$) holds.

So we get *Newton's* formula

$$f(x) = f(\beta_1) + \Delta_1^1 u + \frac{1}{2!}\Delta_1^2 u(u-1) + \ldots + \frac{1}{n!}\Delta_1^n \,.\, u(u-1)\ldots(u-n+1)$$

$$= f(\beta_1) + \frac{\Delta_1^1}{\Delta_x}(x-\beta_1) + \frac{1}{2!}\frac{\Delta_1^2}{\Delta_x^2}(x-\beta_1)(x-\beta_2) + \ldots + \frac{1}{n!}\frac{\Delta_1^n}{\Delta_x^n}(x-\beta_1)\ldots(x-\beta_n).$$

The elements Δ_1^1 can be calculated very easily by the following scheme:—

$$\begin{array}{lllll}
\lambda_1 & & & & \\
 & \Delta_1^1 & & & \\
\lambda_2 & & \Delta_1^2 & & \\
 & \Delta_2^1 & & & \\
\lambda_3 & \cdot & & \Delta_1^3 & \\
\cdot & \cdot & \Delta_{n-2}^2 & & \\
\cdot & \Delta_{n-1}^1 & & & \\
\lambda_n & & & &
\end{array}$$

The degrees of $f(x)$, $\Delta f(x)$, ... , $\Delta^n f(x)$ are decreasing, and the last one is a constant so we can use the above scheme also for *extrapolation* to get the value of $f(x)$ for every arbitrary integral value of u, that means for every value $x=\beta_1+k\Delta_x$, where k is an arbitrary integral number.

Example. Let $f(x)$ be of degree 4, and let $f(1)=3$, $f(2)=4$, $f(3)=5$, $f(4)=1$, $f(5)=2$. In order to get $f(6)$ we use the scheme, calculating at first the numbers above the dotted line from the left to the right, and then the numbers below the dotted line from the right to the left.

$$\begin{array}{lllllll}
1 & 3 & & & & & \\
 & & 1 & & & & \\
2 & 4 & & 0 & & & \\
 & & 1 & & -5 & & \\
3 & 5 & & -5 & & 15 & \\
 & & -4 & & 10 & \cdot\cdot & \\
4 & 1 & & 5 & \cdot & 15 & \\
 & & 1 & \cdot & 25 & & \\
5 & 2 & \cdot & 30 & & & \\
\cdot & \cdot & \cdot & 31 & & & \\
6 & 33 & & & & &
\end{array}$$

Hence $f(6)=33$.

PART V.

MATRICES. RESULTANTS.

§ 1. Matrices.

In the first part of these lectures matrices have been used to solve systems of linear equations, and in § 14 and § 15 a few properties of matrices have been discussed. We will now consider the matrices in a more systematic manner. [1/1]

Let K be an arbitrary field (see Part II § 2), let 0 be the nullelement, 1 be the unitelement of K, and let

$$a,\ b,\ c,\ d,\ e,\ f, \text{ with and without indices of any kind} \qquad (1,\ 1)$$

be arbitrary elements of K. A scheme of m rows and n columns *

$$\begin{pmatrix} a_1^1 & \dots & a_n^1 \\ \dots & \dots & \dots \\ a_1^m & \dots & a_n^m \end{pmatrix} = ((a_k^i)) = \mathrm{A} \qquad (1,\ 2)$$

has been called (see Part I § 6) a *matrix* and a_k^i its *elements*. If $m=n$, this number will be said to be the *degree* of A ; the general case can be reduced to the case $m=n$.

Let A, B, C, ... be matrices of degree n, and let the elements be denoted by the corresponding small latin type, as in (1, 2). The *addition* of matrices is given by

$$\mathrm{A}+\mathrm{B}=\mathrm{F}, \text{ where} \qquad (1,\ 3)$$

$$a_k^i + b_k^i = f_k^i \qquad \text{for } i=1,\dots,\ n,\ k=1,\dots,\ n.$$

The commutative law
$$\mathrm{A}+\mathrm{B}=\mathrm{B}+\mathrm{A} \qquad (1,\ 4)$$

and the associative law
$$(\mathrm{A}+\mathrm{B})+\mathrm{C}=\mathrm{A}+(\mathrm{B}+\mathrm{C}) \qquad (1,\ 5)$$

hold for this addition of matrices.

Let c be an arbitrary element of the field K ; then we define the product

$$c\ \mathrm{A} = ((c\, a_k^i)) \qquad (1,\ 6)$$

i. e., we multiply every element of A with c, and get a new matrix c A of degree n. Then the

distributive laws
$$\begin{aligned} (c+d)\ \mathrm{A} &= c\ \mathrm{A} + d\ \mathrm{A} \\ c\ (\mathrm{A}+\mathrm{B}) &= c\ \mathrm{A} + c\ \mathrm{B} \end{aligned} \qquad (1,\ 7)$$

* Instead of brackets sometimes vertical double bars are put.

hold. For $c=0$, we get

$$0\,A=O, \tag{1, 8}$$

the coefficients of O being equal to 0, and

$$A+O=A. \tag{1, 9}$$

If we define the *subtraction* by

$$A-B=A+(-1)\,B; \tag{1, 10}$$

the matrices of degree n form a *modul* (see Part II [1/2], M in which the matrix O is the nullelement. In M the multiplication with an element of K is a *distributive operation* (see Part I [1/5]).

Hence M is a *modul over* K (see Part II [5/4]). M will be proved to be generated by a set of n^2 independent matrices $E(r,s)$. Let $E(r,s)$ be defined by

$$e_i^r\,(r,s)=1 \qquad e_i^k\,(r,s)=0 \quad \text{for } (i,k)\neq(r,s). \tag{1, 11}$$

then

$$\sum_{i,k} a_i^k\,E(i,k)=A \tag{1, 12}$$

holds, and A is equal to O if and only if every a_i^k is equal to 0. Hence the n^2 matrices are independent. They form a *basis* of M, and M can be considered as a *vectorspace* of rank n^2, (see Part I § 4, Part II [5/4]). By the number n and the field K, the modul m becomes uniquely defined up to an arbitrary isomorphism.

[1/2] In Part I § 14 the multiplication of matrices of M has been defined by

$$A\,B=D, \quad d_i^k=\sum_j a_j^k b_i^j \tag{1, 13}$$

and the associative law $(AB)\,C=A\,(B\,C)$ (1, 14)

has also been proved. From (1, 3) and (1, 13) the

distributive laws $(A+B)\,C=A\,C+B\,C$

$$C\,(A+B)=C\,A+C\,B \tag{1, 15}$$

follow directly.

M is therefore a *ring* (see Part II [1/7]). The reader may prove as an *exercise* that the ring is non-commutative except when $n=1$. The ring has as the unitelement, the matrix

$$E=(e_i^k), \text{ where } e_i^i=1, \text{ and } e_i^k=0 \text{ for } i\neq k. \tag{1,16}$$

Then $$E\,A = A\,E = A. \tag{1, 17}$$

and $$O\,A = A\,O = O \tag{1, 18}$$

for every matrix A. But, in special cases A B may be equal to A, if $B \neq E$, or it may be equal to O although $A \neq O$, $B \neq O$.

E.g., let $A = \begin{pmatrix} a & 0 \\ 0 & 0 \end{pmatrix}$ and $B = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ or $\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$.

To every matrix A there corresponds a determinant det A, and (see Part I § 15)

$$\det(AB) = \det A \cdot \det B. \tag{1, 19}$$

On the other hand, to every element a of K there corresponds a matrix A, for which $\det A = a$ holds. *E.g.*, if $a_1^1 = a$, $a_i^i = 1$ for $i > 1$, and the other elements are equal to 0, this condition holds. Hence in the representation $A \longrightarrow \det A$ the set of the matrices is represented by the *abelian group* (see Part II [1/2]) of the determinants, in such a manner that the multiplication remains invariant.*

$$cA = cEA = \begin{pmatrix} c & & \\ & \ddots & \\ & & c \end{pmatrix} A = A \begin{pmatrix} c & & \\ & \ddots & \\ & & c \end{pmatrix}. \tag{1, 20}$$

It is therefore possible to replace the multiplication of A with an arbitrary element c of K by the multiplication of A with the matrix cE. It may be mentioned that $\det(cE) = c^n$, and therefore

$$\det(cA) = c^n \det A.$$

cE is commutative with every matrix. Hence

$$bBcC = bc\,BC. \tag{1, 21}$$

If $\det A = 0$, then $\det AB = 0 = \det BA$ for every A. Hence the equations $AX = E$, $YA = E$ have no solutions in this case. Let $\det A = a \neq 0$, and let

$$L_i^k = ab_i^k \tag{1, 22}$$

be the cofactor of a_k^i (see Part I p. 22),

* The addition is not invariant, as $\det(A + B) = \det A + \det B$ is in general *not true*.

then $\Sigma a_j^i b_i^l = 0 = \Sigma a_i^l b_j^i$ for $i \neq k$

$\Sigma a_j^i b_i^l = 1 = \Sigma a_i^l b_j^i$ holds [see Part I (34) and (34′)]. Hence

$$AB = E = BA \tag{1, 23}$$

holds. The matrix B, the elements of which have been defined by (1, 22) is said to be the *inverse* of A and will be denoted by A^{-1}.

$$\text{So} \quad AA^{-1} = E = A^{-1}A. \tag{1, 23′}$$

There exist therefore an inverse matrix if and only if the determinant is different from 0.

Let $\det A \neq 0$. then from

$$AX = B \,, \qquad YA = B \quad \text{it follows that}$$

$$X = A^{-1}B, \qquad Y = BA^{-1}.$$

Hence the matrices with non-vanishing determinant form a set in which the two inverse operations of the multiplication can be carried out and give a unique result. This set is not a ring, as the determinant of the sum of two matrices with non-vanishing determinants may be equal to 0. Of course every matrix is the sum of two matrices with non-vanishing determinants; the reader may prove this proposition as an *exercise*.

[1/3] We have to consider now individual matrices. To every matrix A there exists a linear transformation of the space of the n-vectors over K, transforming the n-vector $(x_1, \ldots, x_n)$ into $(x'_1, \ldots, x'_n)$ by

$$a_1^i x_1 + a_2^i x_2 + \ldots + a_n^i x_n = x'_i \tag{1, 24}$$

$$i = 1, \ldots, n.$$

By this transformation the unitvectors (see Part I p. 5) e^k become transformed to the n-vector $(a_1^k, a_2^k, \ldots, a_n^k) = (a_k)$ defined by the kth column. If we want to write formula (1, 24) as a matrix formula, it is convenient to introduce a special notation for matrices in which every element outside the first column is equal to 0.

Let

$$\begin{pmatrix} x_1 & 0 & \ldots & 0 \\ x_2 & 0 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ x_n & 0 & \ldots & 0 \end{pmatrix} = (x) \tag{1, 25}$$

then (1,24) becomes

$$A(x) = (x'). \qquad (1, 24')$$

$$(x) = A^{-1}(x').$$

The formulae (1,24) and (1,24') express the transformation by which the basis of the n-vectorspace formed by the unitvectors e^i becomes transformed to the vectors $\Sigma a_i^k e^i$.

$$\text{Let} \quad (x) = B(y), \qquad (x') = B(y'), \text{ and } \det B \neq 0, \qquad (1, 26)$$

then $$AB(y) = B(y')$$

$$B^{-1}AB(y) = (y'). \qquad (1, 27)$$

By (1, 26) a linear (1, 1)-correspondence between the vectors of the vectorspace has been set up. To the unitvectors of the y-system, there correspond the vectors

$$(\beta_i) = (b_i^1, \ldots, b_i^n) \qquad (1, 28)$$

of the x-system. These vectors form a basis and conversely each basis of the vectorspace can be chosen as a set of vectors (β_i) as it defines the columns of a matrix B with non-vanishing determinant. The vectors $(\beta_i) = (x) = B(y)$ become transformed by (1, 27) in the same manner as the unit-vectors become transformed by (1,24). $(\beta_i) \longrightarrow \Sigma a_i^k (\beta_k)$. So (1, 27) give the transformation of the vectorspace if the transformation of an arbitrary basis is given.

The matrix $$B^{-1}A\,B \qquad (1, 29)$$

is said* to be the *transform* of A by B. As

$$E^{-1}\,A\,E = A,\ B\,(B^{-1}\,A\,B)\,B^{-1} = A,\ C^{-1}(B^{-1}\,A\,B)\,C = (BC)^{-1}A(BC)$$

hold, the transforms of a fixed matrix A form a class (see Part II [1/3]), and from (1,19) it follows that the matrices of this class have the same determinant. The transformations generated by different matrices of the same class are isomorphic, they correspond to the different bases of the vectorspace. If we replace the unitvectors by the basis formed by the vectors (β_i), *i. e.*, if we give to the vector defined by (y) the coordinates $(x_1, \ldots, x_n)$ of $(x) = B(y)$, the matrix $C = B^{-1}\,A\,B$ becomes replaced by A. If C is given, we may arrange that A will be reduced to a *normal-form* by a suitable choice of B.

* Some authors denote this matrix as the transform of A by B^{-1}.

§2. Transformation of A into a normal-form.

[2/1] Let A be a matrix with elements of K; let Λ be a suitable algebraic extension of K, and λ be an arbitrary element of Λ. If the unitvector $(1, 0, \ldots, 0)$ becomes transformed by A into $(\lambda, 0, \ldots, 0)$, then $a_1^1 = , a_1^2 = \ldots = a_0^n = 0$. If every unitvector becomes multiplied with an element of Λ only by this transformation, the matrix A is a diagonal-matrix, and conversely a diagonal-matrix transforms the n-vectorspace in such a manner that the unitvectors become multiplied only with an element of Λ. In geometrical language this transformation means that the direction of the unitvectors will be altered at most by its sign. In order to transform A to a diagonal-matrix, we have to find out n linear independent vectors (1, 28), which become transformed by A into $\lambda_i(\beta_i)$. From [1/3] it follows that $B\,A\,B^{-1}$ transforms the unitvectors in the same manner as $B^{-1}(BAB^{-1})\,B=A$ transforms the vectors (β_i). Hence $B\,A\,B^{-1}$ is a diagonal-matrix.

The conditions

$$a_1^k\,b_i^1 + a_2^k\,b_i^2 + \ldots + a_n^k\,b_i^n = \lambda_i\,b_i^k \qquad (2,\ 1)$$

$$\text{for } k=1, \ldots, n, \quad i \text{ fixed}$$

form a system of n homogeneous linear equations with the matrix

$$A - \lambda_i\,E. \qquad (2,\ 2)$$

Therefore there exist a solution $(b_i^1, \ldots, b_i^n)$ if and only if

$$\det(A-\lambda_i\,E) = 0.$$

As $\det(A-xE)=(-1)^n x^n+(-1)^{n-1}\,\Sigma a_i^i x^{n-1}+\ldots-\Sigma A_i^i x+\det A \qquad (2,\ 3)$

is a polynomial $\chi_A(x)$ in $K[x]$, the field Λ will now be supposed to contain the n roots $\lambda_1, \ldots, \lambda_n$ of (2,3) and

$$\chi_A(x) = \Pi\,(\lambda_i - x). \qquad (2,\ 4)$$

[2/2] For a more detailed investigation of an arbitrary single matrix and its properties, we have to consider integral functions of a matrix. The powers of a matrix A will be defined by the help of the multiplication in the usual manner

$$A^0 = E,\ A^1 = A,\ A^2 = A.A,\ \ldots,\ A^{r+1} = A^r A, \qquad (2,\ 5)$$

and if $\det A \neq 0$, $A^{-m} = (A^{-1})^m$, then it follows that

$$A^s A^t = A^{s+t} = A^t A^s. \tag{2, 6}$$

The powers of A are *commutative* matrices.

Let $\phi(x) = \sum_{i=0}^{s} a^i x^i$ be an arbitrary polynomial in $\Lambda[x]$, then we denote by $\phi(A)$ the matrix

$$\phi(A) = \sum_{i=0}^{s} a_i A^i = a_0 E + a_1 A + \ldots + a_s A^s, \tag{2, 7}$$

and ϕ will be said to be an integral *function* of the matrices over Λ. Let $\psi(x)$ be a polynomial in $\Lambda[x]$, $\omega(x) = \phi(x)\,\psi(x)$. From (2, 6) and (1,21) it follows that

$$\phi(A)\,\psi(A) = \omega(A) = \psi(A)\,\phi(A). \tag{2, 8}$$

Hence the integral functions of a fixed matrix A form a *commutative ring*. We cannot expect that it will include the ring of all matrices of degree n over Λ, as this ring is in general non-commutative (see the 1st exercise in [1/2]). In this ring there are exactly n^2 independent matrices, hence the commutative subring cannot contain more than n^2 independent matrices. The matrices

$$E, A, A^2, \ldots, A^{n^2}$$

therefore cannot be independent, and there must exist a polynomial $\omega(x)$ of degree $\leqq n^2$, with the property that

$$\omega(A) = O.$$

So the matrices may be considered as *roots* of polynomials. The polynomial $\chi_A(x)$ is said to be the *characteristic polynomial* of A. To every solution $\lambda_i^{a_i}$ of $\chi_A(x)$ there corresponds at least one solution $(\beta_i) = (b_1^i, \ldots, b_n^i) \neq (0, \ldots, 0)$ of (2,1).

Theorem. If $\lambda_1, \ldots, \lambda_m$ are $m > 1$ different roots of $\chi_A(x)$, and (β_i) is a solution of (2,9) corresponding to λ_i, $i = 1, \ldots, m$, then the vectors $(\beta_1), \ldots, (\beta_m)$ are independent.

Proof. Let the theorem not be true, then we can choose $t \leqq m$ vectors, say $(\beta_1), \ldots, (\beta_t)$ which are dependent, but every smaller subset

of them is independent. Then an equation

$$a_1(\beta_1) + \ldots + a_i(\beta_i) = (0) \qquad (2,9)$$

holds, where $a_1, \ldots, a_i$ are different from 0. If we transform (2,13) by A we get

$$a_1\lambda_1(\beta_1) + \ldots + a_i\lambda_i(\beta_i) = (0).$$

Hence $a_1[\lambda_1 - \lambda_i](\beta_1) + \ldots + a_{i-1}[\lambda_{i-1} - \lambda_i](\beta_{i-1}) = (0).$

The coefficients are all $\neq 0$, hence $(\beta_1), \ldots, (\beta_{i-1})$ are dependent in contradiction to the supposition. Hence the theorem holds.

Corollary. If the roots $\lambda_1, \ldots, \lambda_n$ of $\chi_A(x)$ are all different, A can be transformed into the diagonal-matrix

$$\begin{pmatrix} \lambda_1 & 0 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & \ldots & 0 & \lambda_n \end{pmatrix}. \qquad (2,10)$$

Proof. Let $B = (b_i^k)$, then the columns of B are independent, hence the rank $(B) = n$, *i.e.*, $\det B \neq 0$. By $B A B^{-1}$ the unitvectors become transformed in the same manner as the vectors (β_i) become transformed by A. Hence $B A B^{-1}$ is equal to the matrix (2, 10).

If the roots are not all different, the corollary does not hold in every case. *E.g.*, Let $A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$, then $\chi_A(x) = (1-x)^2$, $\lambda_1 = \lambda_2 = 1$.

$A E = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$ has the rank 1, hence (2,1) has only one solution $(x_1, x_2) = (1, 0)$.

Remark. The ring of the matrices (A) is homomorphic to the ring $\Lambda[x]$ (see Part II [2/2]), but it is not an *s.r.o.f.*, although $\Lambda[x]$ is ; from $\omega(x) = \varphi(x)\psi(x)$, $\omega(A) = O$, $\varphi(A) \neq O$, it does not follow that $\psi(A) = O$. Hence the number of the matrices which are roots of a polynomial may be greater than the degree of the polynomial ; even it may be infinite.

As $(B^{-1}A B)^n = (B^{-1}A B)(B^{-1}A B) \ldots (B^{-1} A B) = B^{-1}A^nB$,

$\Sigma a_i (B^{-1} A B)^i = B^{-1} \Sigma a_i A^i B$ holds.

Hence from $\varphi(A) = O$ it follows that $\varphi(B^{-1} A B) = O$, where B is an arbitrary matrix with non-vanishing determinant. In other words:

Theorem. Transformed matrices satisfy the same equations.

A similar theorem holds for the characteristic polynomial.

Theorem. To transformed matrices there corresponds the same characteristic polynomial. [2/3]

Proof. As $\det B \det B^{-1} = \det (B\ B^{-1}) = 1$,

$\det (B^{-1}\ A\ B) = \det A$ holds.

$B^{-1}(A - xE)B = B^{-1}A\ B - x\ E$; hence

$\det (A - xE) = \det B^{-1}(A - x\ E)B = \det (B^{-1}A\ B - x E)$ holds.

i.e. $\chi_A(x) = \chi_{B^{-1}AB}(x)$.

The two last theorems have a certain connection between them. It will be proved later on that every matrix is a root of its characteristic polynomial. Now we will consider only that case when the roots of the characteristic polynomial are all different.

As $A(B_i) = \lambda_i(\beta_i)$, $(A - \lambda_i E)(\beta_i) = O$. As the n matrices $A - \lambda_i E$ are commutative, $\chi_A(A)(\beta_i) = \Pi(A - \lambda_i E)(\beta_i) = O$ holds.

The vectors (β_i) form a basis of the n-vectorspace. Hence this vectorspace of rank n is transformed by $\chi_A(A)$ into a vectorspace of rank 0.

Hence (see Part I § 14)

$$\chi_A(A) = O. \tag{2,11}$$

Let λ be a multiple root of the characteristic polynomial of A; then it is a multiple root of the same order, say r, of the characteristic polynomial of $B\ A\ B^{-1}$, where B is an arbitrary matrix of degree n with non-vanishing determinant. To λ there corresponds at least one solution (β_1) of (2,1). Let (β_1) be the first column of a matrix B_1, then the unitvector $(e_1) = (1,0,\ldots,0)$ will be transformed by $B_1 A\ B_1^{-1}$ in the same manner as $(\beta_1) =$ becomes transformed by A. Hence $(e_1) \longrightarrow \lambda(e_1)$. [2,4]

Therefore

$$B_1 A\ B_1^{-1} = \begin{Vmatrix} \lambda & * & \ldots & * \\ 0 & & & \\ \vdots & & A' & \\ 0 & & & \end{Vmatrix} \tag{2,12}$$

In this formula the asterisks * denote certain elements which will not be considered particularly, and A' is a matrix of degree $n-1$.

As $\chi_A(x)=\chi_{B_1AB_1^{-1}}(x)=(\lambda-x)\chi_{A'}(x)$ holds, λ is a root of $\chi_{A'}(x)$ of order $r-1$. We therefore can transform A' into

$$B'A'B'^{-1} = \begin{Vmatrix} \lambda & * & \cdots & * \\ 0 & & & \\ \vdots & & A'' & \\ 0 & & & \end{Vmatrix}$$

where A'' is of degree $n-2$, and if

$$B_1 \begin{Vmatrix} 1 & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & B' & \\ 0 & & & \end{Vmatrix} = B_2 , \qquad (2,13)$$

$$B_2AB_2^{-1} = \begin{Vmatrix} \lambda & * & * & \cdots & * \\ 0 & \lambda & * & \cdots & * \\ 0 & 0 & & & \\ \vdots & & & A'' & \\ 0 & 0 & & & \end{Vmatrix} .$$

The first row of B_2 is (β_1) as we see from (2, 13); let (β_2) be the second row. If $r>2$, then λ is a root of $\chi_{A'}(x)$ of order $r-2$, and we can continue the procedure till we get

$$B_rAB_r^{-1} = \begin{Vmatrix} \lambda & * & \cdots & * & * & \cdots & * \\ 0 & \lambda & * & \cdots & * & * & \cdots & * \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & \lambda & * & \cdots & * & \\ 0 & 0 & \cdots & 0 & & & & \\ \cdots & \cdots & \cdots & \cdots & & A^{(r)} & & \\ \cdots & \cdots & \cdots & \cdots & & & & \\ 0 & 0 & \cdots & 0 & & & & \end{Vmatrix} . \qquad (2,14)$$

$A^{(r)}$ is a matrix of degree $n-r$, $\chi_{A^{(r)}}(x)=\chi_A(x):(x-\lambda)^r$ is not divisible by $(x-\lambda)$. The r first columns $(\beta_1), (\beta_2),\ldots,(\beta_r)$ of B_r are transformed by A in the same manner, as the unitvectors $(e_1),\ldots,(e_r)$ are transformed by $B_rAB_r^{-1}$. Hence

$$A(\beta_1)=\lambda(\beta_1)$$
$$A(\beta_2)=\lambda(\beta_2)+c(\beta_1)$$
$$A(\beta_3)=\lambda(\beta_3)+d_1(\beta_1)+d_2(\beta_2)$$
$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$
$$A(\beta_r)=\lambda(\beta_r)+k_1(\beta_1)+\ldots+k_{r-1}(\beta_{r-1}).$$

Therefore

$$(A-\lambda E)(\beta_1)=(0) \qquad (2,15)$$
$$(A-\lambda E)(\beta_2)=c(\beta_1),\ (A-\lambda E)^2(\beta_2)=(0)$$
$$(A-\lambda E)(\beta_3)=d_1(\beta_1)+d_2(\beta_2^3),\ (A-\lambda E)^2(\beta_3)=d_2c(\beta_1),\ (A-\lambda E)^3(\beta_3)=(0)$$
$$\ldots\ldots\ldots\ldots\ldots\ldots$$
$$\ldots\ldots\ldots\ldots\ldots\ldots(A-\lambda E)^r(\beta_r)=(0).$$

Hence for every vector (a) of the vectorspace V generated by $(\beta_1),\ldots,(\beta_r)$

$$(A-\lambda E)^r(a)=(0) \qquad (2,16)$$

holds. The vectors (β_i) are columns of a matrix with non-vanishing determinant ; hence they are independent. Therefore rank (V) = r.

We want to prove now that every vector (γ) for which

$$(A-\lambda E)^q(\gamma)=(0) \qquad (2,17)$$

holds, belongs to V. If there would be such a vector (γ) outside of V, we can choose it so that $(A-\lambda E)(\gamma)$ is a vector of V. *I. e.*

$$A(\gamma)=\lambda(\gamma)+h_1(\beta_1)+\ldots+h_r(\beta_r).$$

As $(\beta_1),\ldots,(\beta_r),(\gamma)$ are supposed to be independent, there exists a matrix C, of which these vectors form the $(r+1)$ first rows.

Then

$$C\,A\,C^{-1}=\begin{Vmatrix} \lambda & * & \cdots & * & \cdots & * \\ 0 & \lambda & * & \cdots & * & \cdots & * \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & \lambda & * & \cdots & * \\ 0 & \cdots & 0 & & & \\ \cdots & \cdots & \cdots & & A^{(r+1)} & \\ 0 & \cdots & 0 & & & \end{Vmatrix},$$

and therefore $\chi_A(x)=(\lambda-x)^{r+1}\chi_{A^{(r+1)}}(x)$ contrary to the supposition that r is the highest exponent of $(\lambda-x)$ in $\chi_A(x)$. The vectorspace (2, 17) is therefore composed of all vectors, which satisfy (2, 17) for any exponent q. As $(A-\lambda E)(a)=(a')$ satisfies (2,17) if (a) satisfies it, $A(a)=(a')+\lambda(a)$ is a vector of V if (a) belongs to V. By these considerations we get the following theorem.

Theorem. If λ is a root of $\chi_A(x)$ of order r, then the vectors satisfying (2, 17) for any q form a vectorspace V of rank r. Each vector (a) of V satisfies (2, 16) and is transformed by A to a vector of V. If the first r columns of B_r form a suitably chosen basis of V, then (2, 14) holds.

[2/5] Let

$$\chi_A(x) = (\lambda_1 - x)^{r_1} \dots (\lambda_m - x)^{r_m}, \tag{2, 18}$$

where $\lambda_1, \dots, \lambda_m$ are different. To every λ_i there corresponds a vectorspace V_i of rank r_i, so that for every vector (a_i) of V_i, the equation $(A - \lambda_i E)^{r_i}(a_i) = (0)$ holds, and (a_i) is transformed by A into a vector of V_i. To prove that the vectorspaces V_i are independent, we have to use the following lemma.

Lemma. The *h. c. f.* $\phi(x)$ of the polynomials $\phi_1(x), \dots, \phi_m(x)$ of $\Lambda[x]$ can be expressed by $\phi(x) = \phi_1(x)\,\psi_1(x) + \dots + \phi_m(x)\,\psi_m(x)$, where $\psi_1, \dots, \psi_m$ are polynomials of $\Lambda[x]$.

Proof. The lemma is true if $m = 1$ and $m = 2$ (see Part II [4/5]), let it be true for $m = q-1$, and therefore the *h. c. f.* of $\phi_1, \dots, \phi_{q-1}$ be $\omega(x) = \phi_1(x)\,\omega_1(x) + \dots + \phi_{q-1}(x)\omega_{q-1}(x)$. As the *h. c. f.* ϕ of $\phi_1, \dots, \phi_q$ is the *h. c. f.* of ω and ϕ_m, there is $\phi(x) = \omega(x)\,\delta(x) + \phi_m(x)\,\psi_m(x) = \Sigma\,\phi_i(x)\,\psi_i(x)$.

The preceding lemma will be applied to the polynomials

$$\phi_i(x) = \chi_A(x) : (x - \lambda_i)^{r_i}.$$

As these polynomials are relatively prime, there exist m polynomials $\eta_i(x)$ satisfying

$$\eta_1(x) + \dots + \eta_m(x) = 1 \tag{2, 19}$$

$\eta_i(x)$ is divisible by $\phi_i(x)$, $i = 1, \dots, m$ and therefore

$$\text{by } (x - \lambda_k)^{r_k} \qquad \text{for } k \neq i.$$

Hence $\eta_i(A)(a_k) = (0)$, and from (2, 19) it follows that $\eta_i(A)(a_i) = (a_i)$. (2, 20)

If m vectors (a_i) of the different vectorspaces V_i satisfy

$$(a_1) + \dots + (a_m) = 0,$$

we get by multiplication with the matrices $\eta_i(A)$ for every i

$$\eta_i(A)(a_i) = (a_i) = (0).$$

Hence the vectorspaces $V_1, \dots, V_m$ are independent.

If $(a_1^1), \ldots, (a_1^{r_1})$ is a basis of V_1, the $n = r_1 + \ldots + r_m$ vectors

$$(a_1^1), \ldots, (a_1^{r_1}), (a_2^1), \ldots \quad \ldots, (a_m^{r_m}) \tag{2, 21}$$

form a basis of the space of the n-vectors. The matrix C whose columns are the vectors (2,21) has therefore a determinant $\neq 0$. The vectorspaces V_i are invariant for the transformation A. Hence the m vectorspaces generated by

$$(e_1), (e_2), \ldots, (e_{r_1})$$

$$(e_{r_1+1}), \ldots, (e_{r_1+r_2})$$

$$\ldots\ldots\ldots\ldots$$

$$(e_{n-r_m+1}), \ldots, (e_n)$$

are invariant for $C A C^{-1}$. This matrix has therefore the form

$$C A C^{-1} = \left\| \begin{array}{ccc} \boxed{A_1} & & \\ & \boxed{A_2} & \\ & & \boxed{A_m} \end{array} \right\| \tag{2, 22}$$

where A_i is a matrix of degree r_i and its characteristic polynomial is equal to $(\lambda_i - x)^{r_i}$. Every element situated outside the squares is equal to 0. As it has been proved in [2/4], the equation $(A - \lambda_i E)^{r_i}(a_i) = O$ holds for every vector (a_i) of V_i. As the vectors (2,21) form a basis of the total vectorspace, every vector (a) is the sum of vectors (a_i). Hence

$$\chi_A(A)(a) = \Pi (A - \lambda_i E)^{r_i}(a) = (0) \text{ holds for every vector } (a).$$

Hence $$\chi_A(A) = O.$$

The essence of these considerations is given by the following theorem.

Theorem. Let (2, 18) be the characteristic polynomial of A, then there exist m independent vectorspaces V_i of rank r_i, $i = 1, \ldots, m$. These vectorspaces are invariant for A. If the columns of C are basis-

vectors (2, 21) of the vectorspaces V_i, the transformed matrix $C\,A\,C^{-1}$ is of the form (2, 22). A is a root of its own characteristic polynomial.

[2/6] If the basis of V_i is replaced by another basis, the matrix A_i of (2, 22) will be transformed accordingly, the matrices $A_{k \neq i}$ remaining unaltered. To put the matrix (2, 22) into a normal-form, it is therefore only necessary to transform the matrices A_i individually. Hence there is no loss of generality if we suppose that the characteristic polynomial of A is equal to

$$\chi_A(x) = (x-\lambda)^k.$$

In this case there exists for every vector (β) an exponent $q \leqq r$, for which

$$(A-\lambda E)^q(\beta) = (0) \qquad (2, 23)$$

holds. The smallest non-negative number for which (2,23) holds will be said to be the *exponent* of (β),

$$q = \exp(\beta). \qquad (2, 23')$$

Let $0 < p < q$, $c \neq 0$, then

$\exp(c\beta) = q$, $\exp[(A-\lambda E)^p(\beta)] = q-p$, and

if $\exp(\beta_1) > \exp(\beta_2)$, then $\exp[(\beta_1)+(\beta_2)] = \exp(\beta_1)$,

if $\exp(\beta_1) = \exp(\beta_2)$, then $\exp(\beta_1)+(\beta_2) \leqq \exp(\beta_1)$.

From these formulas it follows that the vectors for which $\exp.(\beta) \leqq t$ holds, form a vectorspace W_t, and that in the sequence of the vectorspaces

$$W_1, W_2, \ldots, W_r = V \qquad (2, 24)$$

every vectorspace is included in the subsequent space; it may also be identical to it. For every $s < r$ the vectors

$$(A-\lambda E)^s(\beta)$$

form a vectorspace, which is included in W_{r-s}. The meet of this vectorspace and W_s will be denoted by $W_{s,s}$. Hence in the sequence of the vectorspaces

$$W_{s,r-1}, \quad W_{s,r-2}, \ldots, \quad W_{s,1}, \quad W_s$$

every vectorspace is included in the subsequent spaces. Let

$$u_{r-1}, \quad u_{r-2} \quad , \cdots , \quad u_1 \quad , \quad u_0$$

be the rank of the vectorspaces

$$W_{1,r-1}, \quad W_{1,r-2} \, , \cdots , \quad W_{1,1} \, , \quad W_1$$

then there exists a basis

$$(\beta_1^1) \, , \ldots , (\beta_{u_{r-1}}^1) \, , \; (\beta_{u_{r-1}+1}^1) \, , \ldots , \ldots , \; \ldots , (\beta_{u_1}^1) \, , \; (\beta_{u_1+1}^1) \, , \; \ldots , \; (\beta_{u_0}^1)$$

of W_1 with the property that

$$(\beta_1^1), \ldots , (\beta_{u_i}^1)$$

is a basis of $W_{1,r-1}$ for every index i. (2, 25)

Let $u_s \leqq t < u_{s-1}$, then there exists a vector (β_t^{s+1}) satisfying

$$(A-\lambda E)^s \, (\beta_t^{s+1}) = (\beta_t^1) \tag{2, 26}$$

and we define (β_t^k) , for $1 < k \leqq s+1$ by

$$(A-\lambda E)^{s-k+1} \, (\beta_t^{s+1}) = (\beta_t^k). \tag{2, 26'}$$

From (2, 26) it follows that (2, 26') holds for $k = 1$ too.

Let the vectors (β_t^k) be arranged in a triangular scheme

$$(\beta_1^1) \, , \ldots , (\beta_{u_{r-1}}^1), \; (\beta_{u_{r-1}+1}^1) \, , \; \ldots , (\beta_{u_{r-2}}^1) \, , \ldots (\beta_{u_1}^1), (\beta_{u_1+1}^1), \ldots , (\beta_{u_0}^1) \; ;$$

$$(\beta_1^2) \, , \ldots , (\beta_{u_{r-1}}^2) \, , (\beta_{u_{r-1}+1}^2) \, , \ldots , (\beta_{u_{r-2}}^2) \, , \ldots \ldots , (\beta_{u_1}^2) \; ;$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \tag{2, 27}$$

$$(\beta_1^r) \, , \ldots , (\beta_{u_{r-1}}^r).$$

If we multiply a vector of this scheme from the left side with the matrix $(A-\lambda E)$ we get the vector just above it

$$(A-\lambda E) \, (\beta_t^{k+1}) = (\beta_t^k) \; ; \text{ hence}$$

$$A(\beta_t^{k+1}) = \lambda(\beta_t^{k+1}) + (\beta_t^k) \text{ holds.}$$

The vectorspace U_t generated by the vectors of an arbitrary column, say the t^{th}, is therefore invariant under the transformation by A. These vectors

are independent, *viz.*, from

$$(0)=c_1(\beta_1^1)+\ldots+c_s(\beta_1^s) \qquad \text{it follows that}$$

$$(0)=(A-\lambda E)^{s-1}(c_s\lambda_1^s)=c_s(\beta_1^1), \quad \text{hence } c_s=0.$$

The t^{th} column form therefore a basis of U_t, and in terms of this basis, the transformation of U_t is given by the matrix

$$\begin{Vmatrix} \lambda & 1 & 0 & & 0 \\ 0 & \lambda & 1 & & 0 \\ \ldots\ldots\ldots & \ldots & \ldots & & \\ 0 & & & \lambda & 1 \\ 0 & 0 & & 0 & \lambda \end{Vmatrix}, \tag{2, 28}$$

where the diagonal elements are λ, in the adjacent parallel line the elements are 1 and the other elements are 0. The degree of (2, 28) is equal to $s+1$ where s is given by (2,25). To get the normal-form of the transformation A we have to prove that the vectors of (2, 27) are independent and form a basis of V. The vectors of the first row are independent and form a basis of W_1 by definition.

Let $\sum_{i=1}^{n_1} c_i(\beta_1^2)+(\beta^1) = (0)$ where (β^1) is any vector of W_1 then we get by

multiplication with $(A-\lambda E)$ that $\sum_{i=1}^{n_1} c_i(\beta_1^1) = (0)$, but as the vectors (β_1^1) are

independent, $0 = c_1 = \ldots c_{n_1}$ holds, and therefore $(\beta^1) = (0)$.

Hence the vectors of the two first rows are independent. By mathematical induction it follows * that the vectors (2, 27) are independent.

$$(A-\lambda E)(\beta^2) = \sum_{i=1}^{n_1} d_i (\beta_1^1), \text{ as it is a vector of } W_{1,1}.$$

Let $(\beta^2) = \sum_{i=1}^{n_1} d_i (\beta_i) + (\gamma)$, then $(A-\lambda E)(\gamma) = (0)$, hence (γ) is a

vector of W_1, *i.e.*

$$(\gamma) = \sum_{j=0}^{n_0} k_j (\beta_j^1).$$

* The reader may carry out this mathematical induction as an exercise.

(β^2) is therefore dependent on the vectors of the two first rows. Hence these vectors form a basis of W_2. If (β^2) is a vector of $W_{2,1}$, then $(\beta^2) = (A - \lambda E)(\beta^3)$, $(\gamma) = (A - \lambda E)[(\beta^3) - \Sigma d_i \beta_i]$ belongs to $W_{1,1}$, hence $k_{j>u_1} = 0$.

$(A - \lambda E)(\beta^2) = (A - \lambda E)^2(\beta^3)$ belongs to $W_{1,1}$, hence $d_{i>u_2} = 0$.

The vectors $(\beta^1_1), \ldots, (\beta^1_{u_1}), (\beta^2_1), \ldots, (\beta^2_{u_2})$ form therefore a basis of the vectorspace $W_{2,1}$.

By mathematical induction * it follows, that the vectors of the first k rows form a basis of W_k, and if we omit the vectors $(\beta'_{j>u_r})$ we get a basis of $W_{k,1}$. So we get that for $k = r$, the vectors (2,27) form a basis of the total vectorspace V.

We arrange this basis according to the columns, and we transform in such a manner that the vectors of this basis become unitvectors. Then A will be transformed into

$$C A C^{-1} = \begin{Vmatrix} A^1 & & & \\ & A^2 & & \\ & & A' & \\ & & & A^{u_0} \end{Vmatrix}. \qquad (2, 29)$$

The transformation of the subspace U_r is given by A^r. Hence these matrices have the *normal-form* (2, 28). The degree of the matrix A^r is equal to the length of the corresponding column in (2, 27). These degrees form a monotone non-increasing sequence. On the other hand, if a matrix is given in the normal-form (2, 29), (2, 28), the ranks of the vectorspaces W_1, W_r can easily be found out, the rank of W_k being the number of the matrices A^r, of which the degree is not less than k.

Different normal-forms therefore belong to different classes of matrices. In the general case we can transform each matrix A_k of (2, 22) into the normal-form. By these considerations the following theorem has been proved.

* See p. 86, footnote.

Theorem. Let (2, 18) be the characteristic polynomial of an arbitrary matrix A. Then A can be transformed into a *normal-form* (2, 22), the matrices $A_1, \ldots, A_m$ having the form (2, 29) where the degrees of the $A^1, \ldots, A^{u_0}$ form a non-increasing sequence and each of them has the form (2, 28). In this normal-form, only the permutation of the $A_1, \ldots, A_m$ remains arbitrary.

§3. Some properties of the normal-form and of the characteristic polynomial.

By the correspondence between the normal-forms and the classes of matrices stated in the last theorem we have got a complete insight into the transformations of a vectorspace. It is useful to consider some special cases.

[3/1] If there exists in (2,29) a matrix A^1 of degree r, then there is no other matrix A^r in (2, 29); in this case $u_{r-1}=1$, in every other case it is equal to 0. Every vector of exponent r can be chosen as the vector (β_1^r); we get the other vectors of the basis by repeated multiplication with the matrix $(A-\lambda E)$.

The normal form is the same for all these bases. The matrix (2, 22), will therefore not be altered by the transformation with

$$\begin{pmatrix} \gamma_1, & \gamma_2, & \cdots, & \gamma_r \\ & \gamma_1, & \cdots, & \gamma_{r-1} \\ & & \cdots\cdots & \\ & & & \gamma_1 \end{pmatrix} \qquad (3, 1)$$

where $\gamma_1 \neq 0$ and the elements below the diagonal are equal to 0. Hence (2, 22) is commutative to (3, 1), and there is no other matrix commutative to (2,22) then the matrices (3, 1), as every other transformation will not give a basis of the vectorspace corresponding to the normal-form.

If $u_0=r$, each of the matrices A^r is of degree 1. Hence $0=u_{r-1}=\ldots=u_1$. In this case the vectorspace is identical with W_1. There are no spaces $W_{1,\lambda}$; every basis of the vectorspace leads to the normal-form. This normal-form is therefore commutative to every matrix with determinant $\neq 0$; of course it is a diagonal-matrix with the diagonal-elements $=\lambda$. In the general case the investigation of the admissible transformations of the basis—*i.e.*, the investigation of the matrices commutative to the normal-form—is more complicated, and we

will not go into the details here. The more general problem of finding out the matrices commutative to an arbitrary and fixed matrix A, can easily be reduced to that problem ; *viz.*, if A and B are commutative, $C A C^{-1}$ and $C B C^{-1}$ are commutative too.

The matrix A is a root of the characteristic polynomial $\chi_A(x)$, but it may be that A is a root of a polynomial of lower positive degree. [3/2]

Let

$$\chi_A(x)=(A-\lambda_1 E)^{r_1}(A-\lambda_2 E)^{r_2}\dots(A-\lambda_m)^{r_m}.$$

To $(A-\lambda_1 E)^{r_1}$ there corresponds the minor matrix A_1 in the normal-form (2, 22), and this minor matrix has the form (2, 29), composed by diagonal matrices $A^1, \dots, A^{\omega_0}$. Let s_1 be the degree of A^1, then $1 \leqq s_1 \leqq r_1$ holds, and every vector (a_1) of the vectorspace corresponding to A_1 satisfies the condition

$$(A-\lambda_1 E)^{s_1}(a_1)=(0).$$

Corresponding to this definition we define the integers $s_2, \dots, s_m$; then

$$1 \leqq s_k \leqq r_k \qquad \text{and}$$

$$(A-\lambda_k E)^{s_k}(a_k)=(0) \qquad \text{hold,}$$

for every vector (a_k) of the vectorspace corresponding to A_k.

Let $$\psi(x)=(x-\lambda_1)^{s_1}\dots(x-\lambda_m)^{s_m}, \qquad (3, 2)$$

then $$\psi(A)(\beta)=(0)$$

holds for every vector (β) of the basis (2, 27), and therefore it holds for every vector of the vectorspace.

Hence $$\psi(A)=O. \qquad (3, 3)$$

$\psi(x)$ is a factor of $\chi_A(x)$ and is of lower positive degree except in the case when $r_k=s_k$ for every k which means that the normal-form (2, 29) of every k shows only one matrix, *e.g.*, if $r_1=\dots=r_m=1$.

To prove that A is not a root of a polynomial which is not divisible by $\psi(x)$, we consider the case $\chi_A(x)=(A-\lambda E)^r$. Let $\phi(x)=(x-\lambda)^u\omega$, let ω not be divisible by $(x-\lambda)$, and u be smaller than the degree s of A^1. Then $\phi(A)=(A-\lambda E)^u B$, where $\det B \neq 0$ holds. Hence the transformation of the vectorspace generated by B is an automorphism,

and there exists therefore a vector (β) so that exp. $(B(\beta)) = s > u$. Hence $\phi(A)(\beta) \neq (0)$. Therefore $\phi(A) \neq O$.

Let $A = \left\| \begin{matrix} B & \\ & C \end{matrix} \right\|$. From the law of multiplication of matrices it follows directly that $A^2 = \left\| \begin{matrix} B^2 & \\ & C^2 \end{matrix} \right\|, \ldots, A^q = \left\| \begin{matrix} B^q & \\ & C^q \end{matrix} \right\|$, and therefore $\omega(A) = \left\| \begin{matrix} \omega(B) & \\ & \omega(C) \end{matrix} \right\|$ for every polynomial ω.

Hence $\omega(A) = O$ if and only if $\omega(B) = \omega(C) = O$.

The corresponding rule holds if A is composed of more than two diagonal-matrices. If therefore A has the form (2, 29), $\omega(A) = O$ holds if and only if $\omega(A_i) = O$ for $i = 1, \ldots, m$.

Hence $\omega(x)$ must be divisible by every $(x - \lambda_i)^{s_i}$. By these considerations the following theorem has been proved.

Theorem. $\omega(A) = O$ if and only if $\omega(x)$ is divisible by the polynomial $\psi(x)$, which has been defined by (3, 2).

[3/3] By the *characteristic polynomial* of A, the transformation generated by A becomes uniquely defined only if the roots of the polynomial are all different. If there are equal roots, there are different normal-forms and therefore different non-isomorphic transformations corresponding to the same characteristic polynomial.

To get the characteristic polynomial, it is not necessary to put the matrix in the normal-form. As

$$\chi_A(x) = \det(A - xE),$$

the coefficient of x^k in $\chi_A(x)$ is

$$(-1)^k \Sigma A_{n-k,j}, \tag{3, 4}$$

where $A_{n-k,j}$ denote the minors of A with $n-k$ rows which are symmetric to the diagonal of A. The sums are *invariant* to the transformation;

the most important of these invariants correspond to $k=0$ and $k=n-1$

$$\det(A) \quad \text{and} \quad \Sigma a_{ij}. \tag{3, 4'}$$

We want to apply the theory to the linear substitution of a complex variable

$$w=\frac{\alpha z+\beta}{\gamma z+\delta}, \quad \text{where } \alpha\delta-\beta\gamma\neq 0.$$

We introduce homogeneous coordinates $w=w_1:w_2$, $z=z_1:z_2$;

$$w_1=\alpha z_1+\beta z_2$$

$$w_2=\gamma z_1+\delta z_2.$$

As a common complex factor of α, β, γ, δ is arbitrary, we can arrange that $\alpha\delta-\beta\gamma=1$; now only a common factor ± 1 is arbitrary.

Let $\begin{pmatrix}\alpha & \beta\\ \gamma & \delta\end{pmatrix}=A$. $\chi_A(x)=x^2-\kappa x+1=(\lambda_1-x)(\lambda_2-x)$.

Hence $\lambda_1\lambda_2=1$, $\lambda_1=re^{i\varphi}$, $\lambda_2=r^{-1}e^{-i\varphi}$.

(1) $\lambda_1\neq\lambda_2$, normal-form $\begin{pmatrix}re^{i\varphi} & 0\\ 0 & r^{-1}e^{-i\varphi}\end{pmatrix}$. (3, 5)

As a factor ± 1 and a permutation of λ_1,λ_2 remain arbitrary, we can choose $1\leqq r$, $0\leqq\varphi<\pi$.

(2) $\lambda_1=\lambda_2=\pm 1$, $\kappa=\pm 2$; by a suitable choice of the common factor $=\pm 1$ we can arrange that $\kappa=2$, hence $\lambda_1=\lambda_2=1$.

There are two normal-forms

$$\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix} \quad \text{and} \quad \begin{pmatrix}1 & 1\\ 0 & 1\end{pmatrix}. \tag{3, 6}$$

These transformations are largely discussed in the elements of the theory of functions. The classes of transformations with the normal-form (3, 5) are said to be *loxodromic*, especially for $r=1$ they are called *elliptic*; for $r>1$, $\varphi=0$ they are called *hyperbolic*. The first matrix (3, 6) denotes the identity, the second matrix denotes a *parallel-displacement*, the infinite point of the complex sphere being the only fixpoint, and the other transformations of this class are called *parabolic* transformations, the only fixpoint being a finite point.

§ 4. Theory of elementary divisors.

[4/1] Let S be an *s.r.e.f.* with the following properties.

1. To every element $a \neq 0$ of S there corresponds an integral number $N(a) \geqq 0$. (4, 1)

2. If b is a factor of a, then

$$N(a) \geqq N(b), \tag{4, 2}$$

where the equality holds if and only if a and b are associated.

3. If a_2 is an arbitrary element of S, and a_1 is not divisible by a_2 then there exist in S elements b and a_3 satisfying

$$a_1 + b\, a_2 = a_3,\ N(a_3) < N(a_2). \tag{4, 3}$$

Rings of this kind have been considered in Part II [4/4] and [4/5] of these lectures. Instances of these rings are

(1) The ring of the integral numbers, if we define N for $a \neq 0$ by $N(a) = |a|$. The unities of this ring are $+1$ and -1.

(2) The ring of the integral complex numbers $a + b\,i$ (see Part II § 11), where a and b are integers. We define N for $a + bi \neq 0$ by $N(a+bi) = a^2 + b^2$. The unities of this ring are $+1,\ -1,\ +i,\ -i$.

(3) The ring $K[x]$ of the polynomials in an arbitrary indefinite x over an arbitrary field K. We define N for every polynomial $f(x)$ which is different from the polynomial 0, by $N(f(x)) =$ degree $f(x) + 1$. The unities of this ring are the elements $\neq 0$ of K.

It has been proved in Part II §4: *The factorisation in S is unique and the h.c.f. of two elements a and b can be expressed by*

$$(a, b) = c\, a + d\, b \tag{4, 4}$$

where c, and d are elements of S.

As we see from the lemma of [2/5] the *h.c.f.* of n elements of S can be expressed by

$$(a_1, \ldots, a_n) = c_1 a_1 + c_2 a_2 + \ldots + c_n a_n. \tag{4, 4'}$$

Exercise. Given an *s.r.e.f.* S for which (4, 1), (4, 2), and (4, 3) holds; prove that it is possible to replace the function N by a function N' satisfying the same conditions as N, so that $N'(u) = 1$ if and only if u is a unity of S.

We now shall use the method of "sweep out" (see Part I § 6) on matrices with elements from S. This method needs some modification owing to the fact, that in Part I the elements of the matrices were assumed to be real numbers (or more generally: elements of a field); so the operation of "division by an element" was always possible. In this case we have to deal with the elements taken from a ring; the "division" has therefore to be replaced by the algorithmus of the *h.c.f.* The single steps of the "sweep out" offer two different aspects. They can be considered as operations with rows and columns, and as multiplication (from the left side or from the right side) by certain matrices. It is useful to consider both interpretations at any step. The matrices we have to consider are the same as in Part I § 15. [4/2]

Diagonal-matrices $D = (d_i^k)$, where $d_i^i = d_i$, and

$$\text{for } i \neq k,\ d_i^k = 0; \tag{4, 5}$$

Elementary-matrices $E_{r,s}(\lambda) = ((e_i^k))$, where $r \neq s$

$$e_i^i = 1,$$

$$e_r^s = \lambda, \text{ and} \tag{4, 6}$$

$$\text{every other } e_i^k = 0.$$

We get

DA by multiplying every row of A by the corresponding element d_i

AD „ „ „ „ column A „ „ „ „ „ „ „ „ d_i

$E_{r,s}(\lambda)A$ by row-addition replacing the row (a^r) by $(a^r)+\lambda(a^s)$

$AE_{r,s}(\lambda)$ by column-addition replacing the column (a_s) by $(a_s)+\lambda(a_r)$.

We consider especially those diagonal-matrices

$$U = ((u_i^k)) \tag{4, 7}$$

for which the diagonal-elements $u_i^i = u_i$ are *unities*. By replacing each u_i by u_i^{-1} we get the matrix U^{-1}. As u_i are supposed to be unities, U^{-1} too is a matrix of the type (4,7).

$$E_{r,s}^{-1}(\lambda) = E_{r,s}(-\lambda)$$

is an elementary matrix. Hence if we get B by multiplying A (from the left as well as from the right side) by matrices (4, 6) and (4, 7), we conversely get A by multiplying B by matrices of the same type. B is said to be *congruent to* A, and is denoted as

$$A \underset{\sim}{\cap} B. \tag{4, 8}$$

From $A \subseteq B$, $B \subseteq C$ it follows that $A \subseteq C$, $B \subseteq A$, $A \subseteq A$. Hence the matrices congruent to A form a class, each element of this class being congruent to every other (see Part II [1/3]). To "sweep out A" means to find out a diagonal-matrix congruent to A.

[4/3] Let $a_1, \ldots, a_m$ be m arbitrary but fixed elements of S, and $b_1, \ldots, b_m$ be arbitrary elements of S. The elements

$$a_1 b_1 + \ldots + a_m b_m \tag{4, 9}$$

form a submodul M of S, which is said to be *generated* by $a_1, \ldots, a_m$. Every element of S divisible by any element (4,9) belongs to M; each element of M is divisible by the *h.c.f.* $(a_1, \ldots, a_m)$. But, as it has been shown in [4/1] this *h.c.f.* is an element (4,9). Hence an element of S belongs to M if and only if it is divisible by $(a_1, \ldots, a_m)$. There exists therefore an (1,1)-correspondence between the submoduls M of S and the *h.c.f.* of a finite system generating it, and every M can be generated by the help of only one element.

We have to consider certain submoduls generated by minors of A. Let the determinants

$$A_{k,1}, \ldots, A_{k,m_k} \tag{4, 10}$$

be the minors of degree k of the matrix A, and let the elements of A be elements of the ring S. The minors (4, 10) belong to S, and for every fixed k, (4, 10) generate a submodul of S, say M_k. Let δ_k be the highest common factor of the elements of M_k. In $A_{k,q}$ the cofactors of the elements of an arbitrary row $a'_1, \ldots, a'_k$ are minors of A of degree $k-1$. Hence

$$A_{k,q} = a'_1 A_{k-1,q_1} + \ldots + a'_k A_{k-1,q_k}$$

belongs to M_{k-1}. Hence in the sequence of moduls

$$M_1, \ldots, M_n$$

each modul is a submodul of the preceding, and δ_k is therefore divisible by δ_{k-1}. Hence the *Determinant divisors* δ_k can be represented by the *Elementary divisors* ϵ_k in the following manner

$$\begin{aligned} \delta_1 &= \epsilon_1 \\ \delta_2 &= \epsilon_1 \epsilon_2 \\ &\ldots\ldots \\ \delta_k &= \epsilon_1 \epsilon_2 \ldots \epsilon_k \\ &\ldots\ldots\ldots \\ \det A = \quad \delta_n &= \epsilon_1 \epsilon_2 \ldots \epsilon_n. \end{aligned} \tag{4, 11}$$

Theorem. Congruent matrices have the same elementary divisors.

Proof. Any multiplication from the left side (or from the right side) with a matrix U means that the rows or (the columns) are multiplied with unities. Hence the determinants (4, 10) become multiplied by unities only, and the moduls M_i will remain unaltered. By the row-addition $A \longrightarrow E_{r,s}(\lambda)A$, those minors of A in which the r^{th} row is struck out, will not be altered; the same holds for the minors in which the r^{th} row as well as the s^{th} row occur. Let $A_{i,1}$ be a minor in which the r^{th} row but not the s^{th} occurs, and let $A_{i,2}$ be the minor we get by replacing in $A_{i,1}$ the r^{th} row of A by the s^{th}, then $A_{i,1}$ will be transformed into a sub-modul M_i' of M_i. By the row-addition

$$E_{r,s}(\lambda)A \longrightarrow E_{r,s}(-\lambda)\,(E_{r,s}(\lambda)A) = A.$$

M' will be transformed into a sub-modul M_i'' of M_i'; but as $M_i'' = M_i$, it follows that $M_i = M_i'$. Hence by a row-addition M_i will not be altered, and the same holds obviously for any column-addition. Hence the theorem holds. Especially the *h.c.f.* of the elements of A is $\delta_1 = \epsilon_1$, and is the same for every matrix of the same class.

By the help of multiplication with matrices $E_{r,s}(\lambda)$ and U the following operations can easily be effectuated. [4/4]

(1) Interchanging of two rows (or columns), say (α) and (β).

(α)	(α)	$(\alpha) - ((\alpha)+(\beta)) = (-\beta)$	$(-\beta)$	(β)
(β)	$(\alpha)+(\beta)$	$(\alpha) + (\beta)$	(α)	(α).

(2) "Sweep out" of a row (or a column) by $n-1$ column- additions (or row-additions) if one of the elements, say a is the *h.c.f.* of them

$$a, ab_2, \ldots, ab_n \longrightarrow a, 0, \ldots, 0.$$

(3) If none of the elements of a row $a_1, \ldots, a_n$ is the *h.c.f.* of them we can alter the row by column-addition in such a manner that an element will appear with the property

$$N(b) < N(a_i) \quad \text{for every } i.$$

Proof. Let a_1 be an element for which $N(a_1) \leqq N(a_i)$ holds, and a_2 not be divisible by a_1; then it follows from (4, 3) that there are elements c and b, for which

$$a_2 + ca_1 = b, \quad N(b) < N(a_1) \leqq N(a_i) \qquad (4, 12)$$

holds.

By column-addition we can replace a_2 by b.

(4) If a_1 is the *h.c.f.* of the elements of its row and also the *h.c.f.* of the elements of its column but not of all the elements of the matrix, we can arrange by row- and by column-additions that an element b will appear in the matrix for which $N(b) < N(a_1)$ holds.

Proof. From (1) and (2) it follows that we can suppose a_1 to be the first element of the first row, and the first row and the first column to be "swept out." As by these alterations the *h.c.f.* of the elements of the matrix will not be altered, there exists an element a_2 not divisible by a_1. Let c and b satisfy (4, 12). By interchanging the rows and columns a_2 becomes the 2^{nd} element of the 2^{nd} row. Starting from this position we make row- and column-additions in the following manner.

$$\begin{matrix} a_1 & 0 \\ 0 & a_2 \end{matrix} \longrightarrow \begin{matrix} a_1 & ca_1 \\ 0 & a_2 \end{matrix} \longrightarrow \begin{matrix} a_1 & b \\ 0 & a_2 \end{matrix}.$$

With the help of these operations the "sweep out" can easily be done. Let a be an element of the matrix A for which that function N has a minimum value. If a is different from the *h.c.f.* e_1 of the elements of A, then we can alter A in such a manner that there appears an element b, for which $N(a) \geqslant N(b)$ holds, and if $b \neq e_1$, this procedure can be repeated. But the function N takes integral positive values only, hence after a finite number of steps the procedure must stop, and that is only possible if an element of the matrix becomes equal to e_1.

By (1) e_1 will be placed on the left upper corner of the matrix, and by the help of (2) the first row and the first column will be "swept out." So we get

$$A \sim \begin{pmatrix} e_1 & \\ & A_1 \end{pmatrix}.$$

Each element of A_1 is divisible by e_1; hence the *h.c.f.* is equal to $e_1 e'_2$. By repetition of the procedure we get therefore

$$A \sim \begin{pmatrix} e_1 & \\ & A_2 \end{pmatrix} \sim \dots \sim \begin{pmatrix} e_1 & & \\ & e_1 e'_2 & \\ & & \ddots \\ & & e_1 e'_2 \dots e'_n \end{pmatrix}.$$

The k. c. f. of the minors of degree k is

$$\delta_k = \epsilon_1^{k} \epsilon_2'^{k-1} \dots \epsilon_k'.$$

Hence $$\epsilon_k = \delta_k : \delta_{k-1} = \epsilon_1 \epsilon_2' \dots \epsilon_k'. \qquad (4, 13)$$

$$\delta_k = \epsilon_1 \dots \epsilon_k, \text{ and}$$

$$A \underline{\sim} \begin{pmatrix} \epsilon_1 & & \\ & \epsilon_2 & \\ & & \ddots \\ & & & \epsilon_n \end{pmatrix} \qquad (4, 14)$$

$$\epsilon_k \text{ divisible by } \epsilon_{k-1}. \qquad (4, 15)$$

If on the other hand (4, 14) holds, the determinant divisors δ_k of A have the values given by (4, 13). So we get the following theorem

Theorem. Every matrix A with elements from S is congruent to a diagonal-matrix (4, 14), where ϵ_k satisfy the conditions (4, 15). These elements ϵ_k are identical to the elementary divisors of A. There is therefore an (1, 1) correspondence between the classes of congruent matrices and the sets of elementary divisors, and every set of elements ϵ_k satisfying (4, 15) is an admissible set.

From the preceding theorem it follows that if det A is a unity, A is congruent to a diagonal-matrix U, *viz.*, in this case $\epsilon_1 \dots \epsilon_n$ is a unity and each ϵ_k is a factor of a unity and therefore a unity itself. Hence A is a product of matrices of type (4, 6) and (4, 7) if its determinant is a unity. On the other hand the determinant of a product of matrices (4, 6) and (4, 7) is a unity. The matrices A whose determinants are unities are therefore identical with the products of matrices (4, 6) and (4, 7), and as, to each matrix (4, 6) and (4, 7) there exists an inverse matrix of the same type, every matrix, whose elements belong to S and whose determinant is a unity of S has an inverse of the same type. Therefore the following theorem holds. [4/5]

Theorem. A $\underline{\sim}$ B if and only if

$$A = C_1^{-1} B C_2, \qquad (4, 16)$$

where C_1, C_2 are matrices with elements from S, their determinants being unities of S.

The transformation of a vector space by the matrix A is given by

$$(x) = C_1^{-1} D C_2 (x').$$

Let

$$C_1 (x) = (y), \quad (x) = C_1^{-1}(y)$$
$$C_2(x') = (z), \quad (x') = C_2^{-1}(z),$$

then the modul formed by the vectors (x) is identical with the modul formed by the vectors (y); the modul formed by the vectors (x') is identical with the modul formed by the vectors (z). D is supposed to be a diagonal-matrix with elementary divisors as elements, and

$$(y) = D\ (z) \tag{4, 17}$$

holds.

Exercise. Let (u_1), (u_2), (u_3) be independent vectors in the Euclidean-space, and k_1, k_2, k_3 integral numbers. If the vectors $k_1\ (u_1) + k_2\ (u_2) + k_3\ (u_3)$ start from O the endpoints for all admissible triplets k_1, k_2, k_3 form a *lattice*. Consider the case that S is the ring of the integral numbers and interpret (4, 17) as a general property of lattices.

[4/6] We shall now apply the theory of *congruent* matrices with elements from a ring S as it has been introduced in [4/1] to the theory of the classes of *transformed* matrices with elements from a *field* Λ. The properties of a matrix invariant for transformation and the normal-form got in § 2 will then appear in a new light.

Let Λ be a field containing the coefficients of a fixed matrix A and the root of its characteristic polynomial $\chi_A(x)$. The polynomials $\Lambda[x]$ form an *s.r.o.f.* S satisfying the properties (4, 1), (4, 2), and (4, 3). The elements of Λ are the unities of S. Hence if C is a matrix with elements from Λ and $\det C \neq 0$, then

$$C^{-1} A C \underline{\cup} A, \text{ and}$$

$$C^{-1} A C - xE = C^{-1} (A - x\ E)\ C \underline{\cup} A - x\ E. \tag{4, 18}$$

In order to find out the elementary divisors of $A - x\ E$, we can therefore without any loss of generality suppose that A has the normal-form (2, 22) which consists of a diagonal-system of submatrices $A_1, \ldots, A_m$. Every A_i corresponds to a root λ_i of the characteristic polynomial, and it has the normal-form (2, 29); this means that A_i consists of a diagonal-system of submatrices each of them being of the type (2, 28), *i.e.*, each element of the diagonal is equal to λ_i, the elements just above the diagonal are equal to 1, and all other elements are equal to 0. The reduction of $A - xE$ to a congruent matrix of the type (4, 14) will be effectuated by the help of the

following congruences.

$$\begin{pmatrix} y^r & 1 \\ 0 & y \end{pmatrix} \underline{\sim} \begin{pmatrix} y^r & 1-y^r \\ 0 & y \end{pmatrix} \underline{\sim} \begin{pmatrix} 1 & 1-y^r \\ y & y \end{pmatrix} \underline{\sim} \begin{pmatrix} 1 & 1-y^r \\ 0 & y^{r+1} \end{pmatrix} \underline{\sim} \begin{pmatrix} 1 & 0 \\ 0 & y^{r+1} \end{pmatrix}, \quad (4, 19)$$

and, for $(a, b) = au + bv = 1$,

$$\begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix} \underline{\sim} \begin{pmatrix} a & au \\ 0 & b \end{pmatrix} \underline{\sim} \begin{pmatrix} a & 1 \\ 0 & b \end{pmatrix} \underline{\sim} \begin{pmatrix} a & 1 \\ -ab & 0 \end{pmatrix} \underline{\sim} \begin{pmatrix} 0 & 1 \\ -ab & 0 \end{pmatrix} \underline{\sim} \begin{pmatrix} 1 & 0 \\ 0 & ab \end{pmatrix}. \quad (4, 20)$$

We apply (4, 19) to the sub-matrices of $A_i - xE$, where y is put for $\lambda_i - x$. The elements lying outside of the sub-matrix will not be altered by these row- and column-additions. We start with the two first rows and columns and change them to $\begin{pmatrix} 1 & 0 \\ 0 & (\lambda_i - x)^2 \end{pmatrix}$, the other elements being unaltered. By repetition of this procedure, the different sub-matrices of A_i will be replaced by congruent diagonal-systems with the diagonal-elements.

$$\begin{aligned} &1, \ldots, 1, (\lambda_i - x)^{s_i} \\ &1, \ldots, 1, (\lambda_i - x)^{s'_i} \\ &\cdots\cdots\cdots \\ &1, \ldots, 1, (\lambda_i - x)^{s_i^r} \end{aligned} \quad (4, 21)$$

where $s_i \geqq s'_i \geqq \ldots \geqq s_i^r$, $r_i = s'_i + \ldots + s_i^r$, (4, 22)

and s_i has the same significance as in § 2. Hence $A_i - xE$ is congruent to the diagonal-matrix with the elements (4, 21), and we can arrange them by interchanging the rows and the columns as follows:

$$1, \ldots, 1, (\lambda_i - x)^{s_i}, \ldots, (\lambda_i - x)^{s'_i}, (\lambda_i - x)^{s_i^r}. \quad (4, 23)$$

From (4, 22) it follows that (4, 23) is the normal-form (4, 14) of $A_i - xE$. As the alterations effectuated in the rows and columns of A_i do not alter the other sub-matrices of A, we can reduce these sub-matrices independently to normal-forms (4, 23). So we get a diagonal-matrix congruent to A, the diagonal-elements being powers of $(x - \lambda_i)$ with non-negative exponents. By interchanging the rows and columns and applying the congruence (4, 20) we get a diagonal-matrix

$$1, \ldots, 1, \psi^s(x), \ldots, \psi^1(x), \psi(x), \quad (4, 24)$$

where $\psi(x)=(x-\lambda_1)^{s_1}\ldots\ldots(x-\lambda_m)^{s_m}$ is the same polynomial as it has been defined by (3, 2); every element of (4, 24) is divisible by the preceding elements. Hence (4, 24) is the normal-form (4, 14) of $A-xE$. If (4, 24) is given, we can find out the roots λ_ϱ and the exponents $s_\varrho, s'_\varrho, \ldots, s^{\tau}_\varrho$, and by them the normal-form (2, 22) is uniquely defined. By these considerations we get the connection between the classes of matrices A formed with elements of a *field* K equivalent by transformations and the classes of matrices formed with elements of the *ring* $K[x]$, and congruent to $A-xE$. There is an (1, 1)-correspondence between the classes. The normal-form (4, 24) of $A-xE$ is directly given by the normal-form (2, 22) of A, and conversely we get (2, 22) from (4, 24) by reducing the polynomials $\psi, \psi^1, \ldots$ to products of powers of its linear factors.

§ 5. Hermitian and unitary matrices, Hermitian and quadratic forms.

[5/1] Let K be a sub-field of Λ, and $[\Lambda : K]=2$.

Let α be an element of Λ not belonging to K, then $K(\alpha)=\Lambda$, *viz.*, there is no field including K and included in Λ as $[\Lambda:K]=2$. α is a root of an irreducible polynomial $f(x)$ in $K[x]$ of degree 2; let $\bar{\alpha}$ be the other root. Every automorphism (see Part II [2/2]) of Λ by which the elements of K are not altered, transforms a root of $f(x)$ into a root of $f(x)$. Hence the basis $1, \alpha$ of Λ over K (see Part II p. 32) is either transformed to $1, \alpha$, or to $1, \bar{\alpha}$. In the first case every element of Λ is unaltered by the automorphism; hence the automorphism is the *identity*. In the second case every element $a+b\alpha$ is transformed into $a+b\bar{\alpha}$ for every pair of elements (a, b) of K. This transformation

$$a+b\alpha \longrightarrow a+b\bar{\alpha} \qquad (5, 1)$$

is an automorphism of Λ, *viz.*, the sum (the product) of two elements is transformed into the sum (the product) of the corresponding elements. (5, 1) is the only automorphism of Λ different from the identity by which the elements of K will not be altered. Hence the transformation (5, 1) is independent of the choice of α; elements of Λ corresponding by (5, 1) are said to be *conjugate*. The conjugate of β is denoted by $\bar{\beta}$, then

$$\bar{\bar{\beta}}=\beta \ . \qquad (5, 2)$$

For the purpose of our investigation it is useful to consider [5/2] simultaneously two different cases.

(1) $\Lambda = K$; " conjugate " elements are identical.

(2) $[\Lambda : K] = 2$; " conjugate " elements are defined by (5, 1).

For both cases we make the following condition :

Let (α, β) be an arbitrary pair of elements of Λ, but different from $(0, 0)$, then there exists an element $\gamma \neq 0$, such that

$$\alpha\bar{\alpha} + \beta\bar{\beta} = \gamma\bar{\gamma}. \tag{5, 3}$$

This condition is satisfied *e.g.*

1 if $K = \Lambda$ is the field of the real numbers ;

2 if K is the field of the real numbers, Λ is the field of the complex numbers.

The theory has mostly been applied to these cases, but it is of a more general character, *e.g.*, Λ may supposed to be the field of the algebraic numbers, and K be the field of the real algebraic number, etc. By mathematical induction it follows easily from (5 3) that for every set $(\alpha_1, \ldots, \alpha_n) \neq (0, \ldots, 0)$ there exist an element $\beta \neq 0$ satisfying

$$\alpha_1\bar{\alpha}_1 + \ldots + \alpha_n\bar{\alpha}_n = \beta\bar{\beta}. \tag{5, 3'}$$

As 0 belongs to K, from $\alpha = 0$ it follows that $\bar{\alpha} = 0$; if $\alpha\ \bar{\alpha} = 0$, one of the factors, and therefore both should be equal to 0. The right sides of the equations (5, 3) and (5, 3') are therefore different from 0. Hence, if $(\alpha, \beta) \neq (0,0)$ is any pair of elements of Λ, and if $(a, b) \neq (0,0)$ is any pair of elements of K,

$$\begin{aligned} &\alpha\bar{\alpha} \neq -\beta\bar{\beta}, && a^2 \neq -b \\ \text{especially} \quad &\alpha\bar{\alpha} \neq -1 && a^2 \neq -1. \end{aligned} \tag{5, 4}$$

As the prime field of Λ is a subfield of K, it follows from (5, 4) that the characteristic of Λ is different from 2, and cannot be of the type $4m + 1$. The following theorem will be helpful later on.

Theorem. Let $(\varepsilon_1, \ldots, \varepsilon_n^1)$ be n arbitrary elements of Λ different from $(0, \ldots, 0)$, then there exists a set of n^2 elements ε_k^i of Λ satisfying the

conditions

$$\sum_k u_k^i \bar{u}_k^j = 0 \qquad \text{for } i \neq j \tag{5, 5}$$

$$\sum_k u_k^i \bar{u}_k^i = 1 \tag{5, 6}$$

and

$$\lambda_i u_k^i = v_k^i \qquad \text{for } k = 1, \ldots, n. \tag{5, 7}$$

Proof. Let $x_k = \bar{v}_k^2$ be an arbitrary non-trivial solution of $\sum_k v_k^1 x_k = 0$, then $\sum_k v_k^1 \bar{v}_k^2 = 0 = \sum_k \bar{v}_k^1 v_k^2$ holds. If $n > 2$ there exists a non-trivial solution $y_k = \bar{v}_k^3$ of $\Sigma v_k^1 y_k = \Sigma v_k^2 y_k = 0$; then $\Sigma \bar{v}_k^i v_k^j = 0$ holds, for $i = 1, 2, 3$. $j = 1, 2, 3$, $i \neq j$. We can continue this procedure till we get n^2 elements v_k^i satisfying the conditions

$$(v_1^i, \ldots, v_n^i) \neq (0, \ldots, 0), \ \Sigma v_k^i \bar{v}_k^j = 0, \text{ for } i \neq 0.$$

$\Sigma v_k^i \bar{v}_k^i = \lambda_i \bar{\lambda}_i$ is different from 0. Hence $u_k^i = v_k^i : \lambda_i$ satisfies the conditions (5, 5), (5, 6) and (5, 7).

Exercise. Give an example of fields K, Λ of a characteristic different from 0, for which $[\Lambda : K] = 2$, and (5, 3) holds.

[5/3] The conditions (5, 5), (5, 6) and (5, 7) can be considered as a property of the matrix (u_k^i). To denote properties of this kind in a simple form it is helpful to use the following notations.

Let $A = ((a_k^i))$ be a matrix with elements from Λ, then

$$\overset{*}{A} = ((a_k^{*i})), \text{ for } a_k^{*i} = a_i^k, \text{ is the } \textit{transposed} \text{ of A}, \tag{5, 8}$$

$$\bar{A} = ((\bar{a}_k^i)) \qquad \text{is the } \textit{conjugate} \text{ of A}, \tag{5, 9}$$

$$\overset{\dagger}{A} = (\bar{A})^* = \overline{A^*}. \tag{5, 10}$$

From these formulas it follows that

$$(AB)^* = B^* A^*, \qquad \overline{AB} = \bar{A}\bar{B}, \qquad (AB)^\dagger = B^\dagger A^\dagger,$$

and for $\det A \neq 0$,

$$(A^*)^{-1} = (A^{-1})^*, \quad \bar{A}^{-1} = \overline{A^{-1}}, \quad (A^\dagger)^{-1} = (A^{-1})^\dagger \tag{5, 11}$$

The equations (5, 5) and (5, 6) are therefore equivalent with

$$((u_i^k))^\dagger = ((u_i^k))^{-1}. \tag{5, 12}$$

A matrix satisfying this condition is said to be *unitary*. As the transposed of an unitary matrix is also unitary, a unitary matrix $((u_i^k))$ must also satisfy

$$\sum_k u_i^k \bar{u}_j^k = 0, \qquad \text{for } i \neq j \tag{5, 5'}$$

$$\sum_k u_i^k \bar{u}_j^k = 1. \tag{5, 6'}$$

As $\det((u_i^k)) = \det((u_i^k))^\dagger = \det((u_i^k))^{-1} = 1$: $\det((\bar{u}_i^k))$ holds, hence

$$\det((\bar{u}_i^k)) = \pm 1. \tag{5, 13}$$

A unitary matrix remains unitary after an arbitrary permutation of the rows and of the columns. The product of unitary matrices is unitary.

$$\text{If} \qquad U = \begin{pmatrix} \pm 1 & \\ & \boxed{U'} \end{pmatrix} \tag{5, 14}$$

and one of the matrices U, U' is unitary, the other is unitary too. If the elements of a unitary matrix are elements of K, then $u_i^k = \bar{u}_i^k$, and the matrix is said to be *orthogonal*. This notation corresponds with the notation used in Part I § 9. Using the notion of unitary matrix, we formulate the theorem of [5/2] in the following manner.

Theorem. To every vector $(v) \neq (0)$ there exist unitary matrices with the property that a given row (column) differs from (v) by a factor $\neq 0$ only.

If the element of a matrix H and the roots of its characteristic polynomial* belong to Λ and

$$H = H^\dagger \tag{5, 15}$$

H is said to be an *Hermitian* matrix.

Let H be an Hermitian matrix, U a unitary matrix ; then $H_1 = U^{-1} H U$ has the same characteristic polynomial as H, and $H_1^\dagger = U^\dagger H^\dagger U = H_1$. Hence H_1 is Hermitian. Let λ be an arbitrary root of $\chi_H(x)$, and $(\beta) \neq (0)$

* This condition is essential; it is satisfied always if Λ is a closed field, *e.g.* the field of the complex numbers.

a vector for which $(H-\lambda E)(\beta)=(0)$ holds. Such a vector (β) must exist (see § 2). Let U be a unitary matrix whose first column differs from (β) only by a factor $\neq 0$. By $H_1 = U H U^{-1}$ the first unitvector is transformed in the same manner, as (β) is transformed by H, hence it is multiplied only by λ. The first column of H_1 has therefore the elements $\lambda, 0, \ldots, 0$, and as H_1 is Hermitian,

$$H_1 = \begin{pmatrix} \lambda & & \\ & \boxed{H'} & \\ & & \end{pmatrix}.$$

Hence λ is an element of K. H' is an Hermitian matrix of degree $n-1$. We can transform it by a unitary matrix in the same manner as H has been transformed

$$U'H'U'^{-1} = \begin{pmatrix} \lambda' & & \\ & \boxed{H''} & \\ & & \end{pmatrix}. \quad \text{As the matrix } U_1 = \begin{pmatrix} 1 & & \\ & \boxed{U'} & \\ & & \end{pmatrix} \text{ is}$$

also unitary, $U_1 H_1 U_1^{-1} = \begin{pmatrix} \lambda & & \\ & \lambda' & \\ & & \boxed{H''} \end{pmatrix}$ is an Hermitian matrix.

After n steps we get H transformed into a diagonal-matrix by a matrix which is the product of unitary matrices and is therefore unitary. So we get the following important theorem.

Theorem. An Hermitian matrix H can be transformed by a unitary matrix into a diagonal-matrix, and the roots of $\chi_a(x)$ belong to K.

Let K be the field of the real numbers, Λ be the field of the complex numbers, then it follows from this theorem that the roots of the characteristic polynomial are real. This holds especially, if the matrix is a symmetric real matrix. Hence we can consider any matrix with real elements for which $a_i^k = a_k^i$ holds, as Hermitian, where $K=\Lambda$ is equal to the field of the real numbers, *viz.*, the roots of the characteristic polynomial belong to Λ. Hence we can apply the preceding theorem on this case, the unitary matrices becoming now orthogonal matrices and we get the

Corollary. If in a matrix A with real numbers as elements, $a_i^k = a_k^i$ holds, then $\chi_a(x)$ has real roots only and A can be transformed by an orthogonal matrix with real coefficients into a diagonal-matrix.

The theory of matrices will now be applied to bilinear and quadratic forms. We introduce a double set of indefinites [5/4]

$$x_1, x_2, \ldots, x_n, \ldots, y_1, y_2, \ldots$$
$$\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_n, \ldots, \bar{y}_1, \bar{y}_2, \ldots . \tag{5, 16}$$

Corresponding elements of the two lines will be said to be *conjugate*. They are supposed to be different if $[\Lambda : K] = 2$, and to be identical if $K = \Lambda$. The automorphism of Λ defined in [5/1] by which every element of Λ is replaced by its conjugate, will now be extended to an automorphism of the ring

$$\Sigma = \Lambda [x_1, \ldots, \bar{x}_1, \ldots] \tag{5, 17}$$

of the polynomials in the indefinites (5, 16) with coefficients from Λ. This extension will be made so that every indefinite (5, 16) becomes replaced by its conjugate. So we have simply to replace in every polynomial each coefficient and each indefinite by its conjugate. In the case $K = \Lambda$ the automorphism is the identity. In every case there belongs to every element z of Σ a uniquely defined conjugate element $\bar{z}$, and $\bar{\bar{z}} = z$ holds.

In [1/3] a vector has been represented [see (1, 25)] by a matrix, the first column of which is formed by the coordinates, the elements of the other columns being 0. We will now apply the notations of conjugate matrix and transposed matrix to these special matrices, so that

$$(z) = \begin{pmatrix} z_1 & 0 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ z_n & 0 & \ldots & 0 \end{pmatrix}, \quad (\bar{z}) = \begin{pmatrix} \bar{z}_1 & 0 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ \bar{z}_n & 0 & \ldots & 0 \end{pmatrix}$$

$$(z)^* = \begin{pmatrix} z_1 & \ldots & z_n \\ 0 & \ldots & 0 \\ \ldots & \ldots & \ldots \\ 0 & \ldots & 0 \end{pmatrix}, \quad (z)^\dagger = \begin{pmatrix} \bar{z}_1 & \ldots & \bar{z}_n \\ 0 & \ldots & 0 \\ \ldots & \ldots & \ldots \\ 0 & \ldots & 0 \end{pmatrix}. \tag{5, 18}$$

Let $A = ((a_k^i))$, then $A(y)$ is a matrix with the first element

$$\Sigma a_k^i x_i y_k \tag{5, 19}$$

all other elements being equal to 0. To every matrix A there corresponds

a *bilinear form* (5,19) and conversely. Let

$$(x)=\mathrm{B}(x'), \quad y=\mathrm{C}(y'), \quad \text{and}$$

$$\mathrm{B}^* \mathrm{AC} = \mathrm{A}' = ((a'^i_k)), \tag{5, 20}$$

then $$(x')^* \mathrm{B}^* \mathrm{AC}(y') = (x)^* \mathrm{A}(y).$$

Hence $$\Sigma a^i_k x_i y_k = \Sigma a'^i_k x'_i y'_k. \tag{5, 20'}$$

The formulae (5, 20) and (5, 20') give the transformation of bilinear forms.

We will now consider the case where A is an Hermitian matrix, and where x and y are conjugate ; then B and C are also conjugate ; (5, 19) becomes an *Hermitian form*.

$$\Sigma a^i_k x_i \bar{x}_k, \qquad \text{where} \qquad a^i_k = \bar{a}^k_i, \tag{5, 21}$$

$$\text{and } \Sigma a^i_k x_i \bar{x}_k = \Sigma a'^i_k x'_i \bar{x}'_k, \quad \text{where} \quad ((a'^i_k)) = \mathrm{A}' = \mathrm{C}\dagger \mathrm{AC}. \tag{5, 22}$$

A' is therefore an Hermitian matrix too. In the special case where $\mathrm{K}=\Lambda$, $a^i_k = a^k_i$ and $x_i = \bar{x}_i$; the bilinear form becomes quadratic, and the Hermitian matrix becomes a symmetric one.

$$\Sigma a^i_k x_i x_k = \Sigma a^i_i x_i^2 + 2\sum_{i<k} a^i_k x_i x_k. \tag{5, 23}$$

If the characteristic of K is different from 2, every quadratic form in $x_1, \ldots, x_n$ can be represented in the form (5, 23). We will omit the case of characteristic 2, which needs a special treatment; hence there is an (1,1)-correspondence between the quadratic forms and the symmetric matrices. The transformation is done by

$$((a'^i_k)) = \mathrm{A}' = \mathrm{C}^* \mathrm{AC}, \quad (x) = \mathrm{C}(x'), \quad \Sigma a^i_k x_i x_k = \Sigma a'^i_k x'_i x'_k. \tag{5, 24}$$

This formula for the transformation of quadratic forms reminds us of the transformation of matrices [see (1, 27)], but the notion of inverse matrix has been replaced here by the notion of transposed matrix, and it is not necessary that $\det \mathrm{C} \neq 0$. On applying the theorem of [5/3] and its corollary to (4, 22) and (5, 24) we get the following fundamental theorem.

Theorem. Every Hermitian form can be transformed by a unitary transformation into the normal-form

$$\Sigma a_i x_i \bar{x}_i, \tag{5, 25}$$

and every quadratic form with coefficients from K can be transformed by an orthogonal transformation with coefficients from K into the normal-form

$$\Sigma a_i x_i^2 \qquad (5, 26)$$

a_i being elements of K in both the cases.

Let $K = \Lambda$ be the field of the real numbers, then every quadratic form can be transformed into [5/5]

$$Q(x) = a_1 x_1^2 + \ldots + a_r x_r^2 \quad , \qquad (5, 27)$$

where the coefficients are different from zero, by the transformation with an orthogonal matrix. As the determinant of an orthogonal transformation is equal to ± 1, the rank of the matrix A will not be altered by the transformation, and is therefore equal to r. If we replace x by $-x$ the formula (5, 27) will not be altered, but the sign of the orthogonal transformation becomes changed. Hence we can transform an arbitrary quadratic form into a normal-form (5, 27) by an orthogonal transformation with determinant $+1$. Such a transformation means in the geometry of the space of n dimensions a rotation through the origin. This transformation is well known in the analytic geometry of conics and of quadrics as the *transformation to the principal axes*.

Let $a_i = \pm b_i^2$, and $b_i x_i = y_i$, then (5, 27) is transformed to a sum of squares with certain signs $\pm$. After a permutation of the indices, and replacing y by x we get therefore

$$q(x) = x_1^2 + \ldots + x_p^2 - x_{p+1}^2 - \ldots - x_r^2. \qquad (5, 28)$$

Hence every quadratic form can be transformed to (5, 28) by a non-degenerated linear transformation with real coefficients. The integer r is the rank of matrix of the quadratic form and therefore invariant. We will prove that p is an invariant too.

Theorem. Every quadratic form with real coefficients can be transformed into one and only one normal-form $q(x)$ by a non-degenerated linear transformation with real coefficients.

Proof. As it has been shown above, the transformation into the normal-form is always possible, and r is an invariant. We have therefore only to prove that a transformation of $q(x)$ into

$$q_1(x) = x_1^2 + \ldots + x_{p'}^2 - x_{p'+1}^2 - \ldots - x_r^2$$

by linear non-degenerated substitution is possible only if $p=q$. Let $p \neq q$, say $p \geqq q$ without any loss of generality.

$$x_i = b_1^i z_1 + \ldots + b_r^i z_r$$

$$z_i = c_1^i x_1 + \ldots + c_r^i x_r \qquad i=1, \ldots, r.$$

$q(x)=q_1(z)$ for corresponding systems $(x_1, \ldots, x_r)$ and $(z_1, \ldots, z_r)$. The $q+r-p<r$ linear homogeneous equations

$$c_1^k x_1 + \ldots + c_r^k x_r = 0 \qquad k=1, \ldots, q$$

$$x_t = 0 \qquad t=p+1, \ldots, r$$

have a solution $(\xi_1, \ldots, \xi_p, 0, \ldots, 0)$ different from $(0, \ldots, 0)$ *viz.* the rank of the matrix of this system of equations is $\leqq q+r-p<r$. The corresponding values of $z_1, \ldots, z_r$ are $(0, \ldots, 0, \zeta_{q+1}, \ldots, \zeta_r)$ and different from $(0, \ldots, 0)$. Hence $q(\xi)>0$, $q_1(\zeta)<0$ in contradiction to $q(\xi)=q_1(\zeta)$.

A quadratic form is said to be *positive definite* if $n=r=p$; it is *negative definite* if $n=r$, $p=0$; it is *semidefinite* if $n>r$, $p=r$ or $=0$, and it is *indefinite* if $r>p>0$.

[5/6] Finally we will give a geometrical interpretation of the last results without going into the details.

In the projective $(n-1)$-dimensional space, a quadric becomes represented by

$$\rho \Sigma a_{ik} x_i x_k = 0,$$

where $\rho \neq 0$ is an arbitrary factor.

Hence the quadric has one and only one normal-form

$$\rho q_1(x) = 0,$$

and the sign of ρ can be fixed in such a manner that $q(x)$ has not fewer positive than negative terms. We get therefore the different types of quadrics in the projective $(n-1)$-dimensional space given by the different normal-forms

$$q(x) = 0, \quad \text{for} \quad r=1, \ldots, n, \ {}^{r}/_{2} \leqq p \leqq r. \tag{5, 29}$$

Especially the quadrics without any real point are those for which $p=r=n$. The normal-form (5, 29) has the property that every fundamental point of the coordinate system *i.e.* every point, for which all the coordinates are equal to 0 except x_i, is polar to the opposite hyperplane $x_i=0$.

In the affine n-dimensional space the quadrics are given by

$$\Sigma a_{ik} x_i x_k + \Sigma b_i x_i + c = 0.$$

On applying the theorem of [5/5] we get easily the following types of affine normal-forms.

$$\begin{aligned} &q(x) = 0, && r = 1, \ldots, n && r/2 \leqq p \leqq r \\ &q(x) = 1, && r = 1, \ldots, n, && p \leqq r \\ &q(x) = x_{r+1}, && r = 1, \ldots, n-1, && r/2 \leqq p \leqq n. \end{aligned} \tag{5, 30}$$

If we replace $q(x)$ by the quadratic form $Q(x)$ of (5, 27) we get the types of quadrics different in the sense of metric geometry. The formula (5, 27) can also be interpreted for projective geometry.

Let $R(x_1, \ldots, x_n)$ and $S(x_1, \ldots, x_n)$ be two quadratic forms, S representing a quadric without real points, then we transform the coordinates in such a manner that S is transformed to $S' = x_1^2 + \ldots + x_n^2$, and R to R'. By any orthogonal transformation, S' will not be altered, but we can transform R' by an orthogonal transformation to the normal-form (5, 27). Hence we can transform simultaneously

$$\begin{aligned} &R \quad \text{to} \quad a_1 x_1^2 + \ldots + a_n x_n^2, \\ &S \quad \text{to} \quad x_1^2 + \ldots + x_n^2 \qquad \text{and the pencil} \\ &\kappa R + \lambda S \quad \text{to} \quad (\kappa a_1 + \lambda) x_1^2 + \ldots + (\kappa a_n + \lambda) x_n^2. \end{aligned}$$

The elements of the the pencil have therefore been transformed to the normal-form simultaneously.

§ 6. Resultants.

Let [6/1]

$$f(x) = x^n + a_1 x^{n-1} + \ldots + a_n = (x - \alpha_1). \ldots .(x - \alpha_n) \tag{6, 1}$$

$$g(x) = x^m + b_1 x^{m-1} + \ldots + b_m = (x - \beta_1). \ldots .(x - \beta_m) \tag{6, 2}$$

$$R(f, g) = \prod_i \prod_k (\alpha_i - \beta_k). \tag{6, 3}$$

Then $R(f, g)$ is uniquely defined by the polynomials f and g, and it is said to be the *resultant* of f and g. The necessary and sufficient condition that f and g may have a common root is that the resultant is equal to zero.

From (6, 3) it follows

$$R(f, g) = (-1)^{mn} R(g, f); \tag{6, 4}$$

from (6, 2) and (6, 3)

$$R(f, g) = \prod_{i=1}^{n} g(a_i), \tag{6, 5}$$

and by interchanging f and g, and applying (6, 4) we get

$$R(f, g) = (-1)^{mn} \prod_{k=1}^{m} f(\beta_k). \tag{6, 6}$$

The right side of the equation (6, 5) is a symmetric polynomial in $a_1, \ldots, a_n$ with coefficients $h_i = h_i(b_1, \ldots, b_m)$, these polynomials h_i having integral coefficients. From Part II [10/3] it follows that $R(f, g)$ can be represented as a polynominal in the elementary symmetric polynomials of $a_1, \ldots, a_n$, with coefficients $\phi_k = \phi_k(h_1, \ldots, h_l)$, where ϕ_k has integral coefficients. As the elementary symmetric polynomials of $a_1, \ldots, a_n$ are equal to $\pm a_i$, the resultant $R(f, g)$ can be represented as a polynomial in $a_1, \ldots, a_n, b_1, \ldots, b_m$ with integral coefficients.

For a reason which will become evident later on, this representation will be written as

$$R(f, g) = R(1, a_1, \ldots, a_n; 1, b_1, \ldots, b_m). \tag{6, 7}$$

If in any term $A = a_1^{s_1} a_2^{s_2} \ldots a_n^{s_n} b_1^{t_1} b_2^{t_2} \ldots b_m^{t_m}$ the factors a_i are represented by $\alpha_1, \ldots, \alpha_n$ and the factors b_k by $\beta_1, \ldots, \beta_m$, then A becomes an homogeneous polynomial in α_i and β_i of degree *

$$t = s_1 + 2s_2 + \ldots + ns_n + t_1 + 2t_2 + \ldots + mt_m. \tag{6, 8}$$

t is said to be the *weight* of A. From (6, 3) it follows that $R(f, g)$ is homogeneous of degree $n\,m$; hence each term of $R(1, a_1, \ldots, a_n; 1, b_1, \ldots, b_m)$ has the weight $n\,m$. From (6, 5) we see that one of these terms is equal to b_m^n.

Let S be a polynomial in $a_1, \ldots, a_n, b_1, \ldots, b_m$ with the property, that S becomes equal to zero if f and g have a common root. By representing a_i as symmetric polynomials in $\alpha_1, \ldots, \alpha_n$, and b_k as symmetric polynomials

* See Part II, p. 19.

in $\beta_1,\ldots,\beta_m$, we get

$$S = \Sigma = \Sigma\,(a_1,\ldots,a_n,\ \beta_1,\ldots,\beta_m).$$

The right side of this equation is equal to zero if $a_1 = \beta_1$. Hence

$$\Sigma = \Sigma\,(a_1,\ldots,a_n,\ \beta_1,\ldots,\beta_m) - \Sigma\,(a_1,\ldots,a_n,\ a_1,\ldots,\beta_m).$$

Subtracting the corresponding terms on the right side, we see that Σ is divisible by $a_1-\beta_1$, and in the same manner it follows that Σ is divisible by $a_i-\beta_k$; hence Σ is divisible by $R(f,g)$. Or $R(f,\ g)=(\Sigma,\ R(f,g))$. From the 2nd theorem of Part II [10/5] it follows therefore that $R(1,\ a_1,\ldots,a_n\ ;\ 1,\ b_1,\ldots,b_m) = (S,\ R(1,\ a_1,\ldots,a_n;\ 1,\ b_1,\ldots,b_m))$. Hence S is divisible by the resultant. The weight of every term is therefore not less than $m\ n$; if each term has the weight $m\ n$, S differs from the resultant by a factor of weight 0 only, and this factor is the coefficient of the term $b_m{}^n$ in S.

To get the resultant as a polynomial in $a_1,\ldots,a_n,\ b_1,\ldots,b_m$, we have therefore to find out a polynomial S, with the following three properties: [6/2]

1. S=0 if f and g have a common root.
2. Each term of S has the weight $m\ n$.
3. The term $b_m{}^n$ has the coefficient 1.

A polynomial of this kind can easily be found out by the following consideration.

Let $f(a)=0=g(a)$, then the following $n+m$ equations hold

$$\begin{array}{llr}
a^m f(a) = a^{n+m} + a_1 a^{n+m-1} + \ldots + a_n a^m & & =0 \\
a^{m-1} f(a) = & a^{n+m-1} + \ldots + a_{n-1} a^m + a_n a^{m-1} & =0 \\
\ldots\ldots\ldots & & \\
a\ f(a) = & a^{n+1} + a_1 a^n + \ldots + a_n a & =0 \\
a^n\ g(a) = a^{n+m} + b_1\ a^{n+m-1} + \ldots + b_m a^n & & =0 \\
a^{n-1}\ g(a) = & a^{n+m-1} + \ldots + b_{m-1}\ a^n + b_m\ a^{n-1} & =0 \\
\ldots\ldots\ldots & & \\
a\ g(a) = & a^{m+1} + b_1\ a^m + \ldots + b_m a & =0.
\end{array}$$

We consider this system as a system of linear equations in $a^{n+m-1}, \ldots, a_0$. It can be satisfied only if its determinant is equal to zero. Hence a necessary condition for that f and g may have a common root is

$$S = \begin{vmatrix} 1 & a_1 & \ldots & a_n & & \\ & 1 & & a_{n-1} & a_n & \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ & & 1 & a_1 & \ldots & a_n \\ 1 & b_1 \ldots b_m & & & & \\ & 1 \ldots b_{m-1} & b_m & & & \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ & & & 1 & b_1 \ldots & b_m \end{vmatrix} = 0. \qquad (6, 9)$$

The terms of S are of weight $m\,n$ each, the term $b_m{}^n$ is the diagonal-element and has the coefficient 1. Hence

$$S = R(f, g). \qquad (6, 9')$$

[6/3] Let $F(x) = a_0 x^n + \ldots + a_n$

$G(x) = b_0 x^m + \ldots + b_m$,

where nothing is supposed about the coefficients. We *define* now

$$R(F, G) = R(a_0, \ldots, a_n; b_0, \ldots, b_m)$$

$$= \begin{vmatrix} a_0 & \ldots & a_n & & \\ & \cdots & \cdots & \cdots & \\ & & a_0 & \ldots & a_n \\ b_0 & \ldots & b_m & & \\ & \cdots & \cdots & \cdots & \\ & & & b_0 & \ldots b_m \end{vmatrix} \qquad (6, 10)$$

As in (6, 9) there are m rows with elements a_i and n rows with elements b_i. If $a_0 = b_0 = 1$, then $F = f$, $G = g$, and we see from (6, 9) and (6, 9'), that the notation $R(F, G)$ conforms to $R(f, g)$. We have to consider three cases:

1. $a_0 \neq 0$, $b_0 \neq 0$.

$$F(x) = a_0 \left(x^n + \ldots + \frac{a_n}{a_0} \right) = a_0 \varphi(x) = a_0 (x - \alpha'_1) \ldots (x - \alpha'_n)$$

$$G(x) = b_0 \left(x^m + \ldots + \frac{b_m}{b_0} \right) = b_0 \psi(x) = b_0 (x - \beta'_1) \ldots (x - \beta'_m).$$

From (6,10) it follows that*

$$R(F, G)=a_0^m b_0^n R(\phi, \psi). \qquad (6, 11)$$

Hence for (6, 3) (6, 5), and (6, 6)

$$R(F, G)=a_0^m b_0^n \prod_i \prod_k (\alpha_i-\alpha'_k)=a_0^m \prod_i G(\alpha'_i)=(-1)^{mn} b_0^n \prod_k F(\beta'_k). \qquad (6, 12)$$

Hence in this case $R(F, G)=0$ is the necessary and sufficient condition for the existence of a common root of F and G.

2. $a_0=0$, $b_0=0$. From (6,10) it follows that $R(F, G)=0$, independent of the existence of a common root.

3. $a_0 \neq 0$, $b_0=0$ (or $a_0=0$, $b_0 \neq 0$).

Let $b_1= \dots =b_m=0$; then $R(F, G)=0$, and every root α' of F satisfies obviously $G(\alpha')=0$.

Let every $b_{k<r}=0$, $b_r \neq 0$. $G_i(x)=b_i x^{m-i}+ \dots +b_m$ for $i=1, \dots, r$.

By setting 0 for b_0 in (6,10) we get $R(F, G) = a_0 R(F, G_1)$

and by mathematical induction $R(F, G) = a_0^r R(F, G_r)$.

Hence $R(F, G)=0$ if and only if F and G_r have a common root, *i.e.*, if F and G have a common root. The corresponding proposition holds for $a_0 = 0$, $b_0 \neq 0$. By these considerations we get the following theorem:

Theorem. If $F(x)$ and $G(x)$ have a common root, then $R(F, G)=0$. If $R(F, G)=0$, then either $a_0=b_0=0$, or $F(x)$ and $G(x)$ have a common root.

Exercises. 1. Consider the correctness of (6, 11) and (6, 12) in the case 3.

2. Let $F(x, y)=\Sigma a_i x^{n-i} y^i$, and $G(x, y)=\Sigma b_k x^{m-k} y^k$.

State the necessary and sufficient condition for $(F, G) \neq 0$.

Let $u_1, \dots, u_m, v_1, \dots, v_n$ be the cofactors * of the last column of the determinant (6, 10), and let [6/4]

$$u(x)=u_1 x^{m-1}+ \dots +u_{m-1}x+u_m, \quad v(x)=v_1 x^{n-1}+ \dots +v_{n-1}x+v_n.$$

* See Part I, p. 22.

When we multiply the $n+m$ equations

$$x^{m-1}F(x) = a_0 x^{n+m-1} + \ldots + a_n x^{m-1}$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$F(x) = \qquad\qquad a_0 x^n + \ldots + a_n$$

$$x^{n-1}G(x) = b_0 x^{n+m-1} + \ldots + b_m x^{n-1}$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$G(x) = \qquad\qquad b_0 x^m + \ldots + b_m$$

with $u_1, \ldots, u_m, v_1, \ldots, v_n$ respectively and add, we get

$$u(x)\,F(x) + v(x)\,G(x) = R(F, G). \qquad (6, 13)$$

We can express this formula by the following theorem.

Theorem. Let R be the ring generated by the indefinite x and the coefficients of F and G, then R(F, G) is linearly dependent on F and G with coefficients from R.

Exercise. Prove the theorem of [6/3], without any reference to symmetric functions, by the help of the last theorem. (Special attention should be given to the case when every cofactor is equal to zero.)

If $F(x)$ and $G(x)$ have a common root, the highest common factor (F, G) is a polynomial of positive degree; we can get it by the algorithmus of the *h.c.f.*, hence its coefficients belong to the ring generated by the coefficients of F and G by addition, subtraction and multiplication.

[6/5] Let the coefficients $a_0, \ldots, a_n$; $b_0, \ldots, b_m$ of $F(x)$ and $G(x)$ be polynomials of $K(y)$, K being an arbitrary field;

$$F(x) = f(x, y), \quad G(x) = g(x, y).$$

The resultant R(F, G) is therefore a polynomial in y, and from (6, 13)

$$R(F, G) = R(y) = u(x, y)\,f(x, y) + v(x, y)\,g(x, y). \qquad (6, 14)$$

We suppose that at least one of the polynomials $a_0 = a_0(y)$, $b_0 = b_0(y)$ is different from the polynomial 0. From [6/3] it follows that $f(x, y)$ and $g(x, y)$ have a common factor depending on x, if and only if $R(y)$ is the polynomial 0. The procedure of getting $R(y)$ and $g(x, y)$ is called *elimination of* x. Let η be a root of $R(y)$; then $R(f(x, \eta), g(x, \eta)) = R(\eta) = 0$. Hence either $a_0(\eta) = b_0(\eta) = 0$, or $f(x, \eta)$, $g(x, \eta)$ have a common root. By this method we can find out the common solutions of two equations

$$f(x, y) = 0, \qquad g(x, y) = 0.$$

CORRECTIONS

Part I. (see the corrections given in Part II.)

Page.	*Line.*	*Read*	*For*
iii (Preface)	19	does	do
13	19	(2/H)	(2)
16	14	depend on	apply to
20	10	10.	(11)
	23	11.	10.

Part II.

5	10	$c+O$	$c+\bar{O}$
6	19	M	M'
11	22	$(b'+w)$	$(b'+t)$
	23	$a'w$	$a't$
13	20	are	and also the distributive law are
	22	a non-distributive system	a system
	28	the reader may verify that (2, 8) generates an addition and a multiplication of the classes for which the commutative, associative, *and* distributive laws hold, and	we should only prove that
13	21	field	fleld
15	11	a_{n-1}	a_{n-1}
21	12 and 13 : interchange the exponents !		
23	24	$f(x)$, $\phi(x)$, $\psi(x)$	$f(x)$, $\psi(x)$, $\psi(x)$
26	10	$\frac{a_n}{b_m}$	$\frac{b_m}{a_n}$
	11	$\frac{a_{n_i}}{b_m}$	$\frac{b_m}{a_{n_i}}$
	26	$\phi(x)\,\phi_1(\chi)$	$\phi(x)\psi_1(x)$
33	6	K	K_1^-
	20	K(β)	K(α)
38	9	N	K [three times!]

Page.	*Line.*	*Read*	*For.*
42	18	to K(a)	to K
46	9	a_i	a_i
47	19	$x^{n-1} + \ldots + b_n$	$x^{m-1} + \ldots + b_m$
	30	$F(a_1, \ldots, a_n)$	$F(a_1, \ldots, a_n)$
52	17	$\bar{h}$	h [error not in all copies]
61	3	(13, 2)	(13, 2
	5	(13, 2)	(13, 1)
	6	$u(n-1)$	$n(n-1)$

Parts III—V. (in this volume)

3	19	P′	P [twice]
	20	Q′	Q [twice]
	25	a_{i+i+1}	a_{i+i-1}
8	25	$a-s$	$s+1-a$
11	5	$>a>$	$<a<$
12	4	+	− [between the fractions]
13	27	$\frac{P_{2n}}{Q_{2n}}$	$\frac{P_{2n}}{P_{2n}}$
14	17	$s_n \mid \lambda$	s_n, λ
16	1	a	a
	15, 19	purely periodic	purely
19	3	[3/4] [on the margin]	
	9	>1	>0
20	12	$a+\beta$	$a_1+\beta_1$
		$s+a$	$s+a_1$ [twice]
22	5	$\sqrt{26}$	26
31	1	Σs_i	s_i
	6	If this sum is convergent, the sum taken for odd indices is divergent, where it follows easily that $Q_{2n+1} \to \infty$. Hence	Hence
	10	interchange "then" with "and"	

Page.	*Line.*	*Read*	*For*
41	17	(1, 3)	(3)
42	21	$(x-a)(x-\bar{a})$	$(x-\bar{a})(x-\bar{a})$
53	16	(1, 30)	(30)
58	5	2·40824	2·40224
59	15	$\left(1+\frac{2b_3^m+b_4^m}{b_1^m}\right)$	$\frac{(1+2b_3^m+b_4^m)}{b_1^m}$
61	14	b_1	b_6
65	15	[5/2]	[5/]
66	22	Δ_{k-i}	$\Delta_{k\,i}$
72	7	(see Part II [1/2])	(see Part II [1/2],
	18	modul M	modul m
	30	$((e_k^i))$, where $e_i^i=1$, and e_k^i	(e_k^i), where $a_i^i = 1$, and a_k^i
74	7	exists	exist
75	18	a_k^i	a_k^i
76	21	exists	exist
79	26	[omit=]	
80	10	[place λ from the 2nd to the 1st column in the 1st line of (2, 14)]	
	11	$A^{(r)}$	$A^{(n)}$
		$\chi_{A^{(r)}}$	$\chi_{A^{(s)}}$
82	14	[replace the index m by q)	
85	9	$W_{1,\,i}$	$W_{1,\,r-1}$
101	17	exists	exist
105	23	$(x)^*A(y)$	$A(y)$